PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/81, 15/55, 15/56, 15/11, 15/62, 1/19 // (C12N 1/19, C12R 1:645)

A1

(11) International Publication Number: WO 94/13821

(43) International Publication Date: 23 June 1994 (23.06.94)

(21) International Application Number: PCT/EP93/03547

(22) International Filing Date: 9 December 1993 (09.12.93)

(30) Priority Data:
92203932.6 11 December 1992 (11.12.92) EP
(34) Countries for which the regional or international application was filed: NL et al.
93200240.5 29 January 1993 (29.01.93) EP
(34) Countries for which the regional or international application was filed: NL et al.

(71) Applicant *(for all designated States except US)*: QUEST INTERNATIONAL B.V. [NL/NL]; Huizerstraatweg 28, NL-1411 GP Naarden (NL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: CHAPMAN, John, William [GB/NL]; Weena 931, NL-3013 AL Rotterdam (NL). MUSTERS, Wouter [NL/NL]; Wipperspark 138, NL-3141 RD Maassluis (NL). ROUWENHORST, Robert, Jan [NL/NL]; Oostsingel 95, NL-2612 HG Delft (NL). TOSCHKA, Holger, York [DE/DE]; Langnese-Iglo GmbH, Ackern 1, D-4421 Reken (DE). VERBAKEL, Johannes, Maria, A. [NL/NL]; Ingeland 9, NL-3155 GC Maasland (NL).

(74) Agent: UNILEVER N.V.; Patent Division, P.O. Box 137, NL-3130 AC Vlaardingen (NL).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, LV, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: THE USE OF THE *KLUYVEROMYCES MARXIANUS* INULINASE GENE PROMOTER FOR PROTEIN PRODUCTION

(57) Abstract

The invention provides a nucleic acid sequence derivable from a yeast, e.g. from a *Kluyveromyces* yeast, preferably from the strain *K. marxianus* var. *marxianus* (CBS 6556), and comprising at least a regulatory region derivable from a gene encoding a polypeptide having inulinase activity or a functional modification thereof. Regulatory regions comprise a promoter, a terminator, and a sequence encoding a secretory signal necessary for secreting a gene product from a yeast. The DNA sequence of the inulinase gene from *K. marxianus* var. *marxianus* (CBS 6556) is given, including its regulatory regions. The regulatory regions can be used for preparing vectors suitable for transforming a host, preferably a yeast, for the production of desired expression products, e.g. proteins, RNA suitable for flavouring purposes, and metabolites. A process for production of an expression product is also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

The use of the *Kluyveromyces marxianus* Inulinase gene promoter for protein production.

Technical Field

The subject invention lies in the field of DNA technology. In particular the invention covers a nucleic acid sequence derivable from a yeast and comprising at least a regulatory region derivable from a gene encoding a polypeptide having inulinase activity. The invention is also directed at an expression vector comprising the aforementioned nucleic acid sequence and is furthermore directed at the use of said nucleic acid sequence or expression vector for producing a desired expression product.

Background

Yeast strains of the genus *Kluyveromyces* have been used for the production of enzymes for many years, and the growth of these strains has been extensively studied. *Kluyveromyces marxianus* var. *marxianus* strains (hereinafter also called *Kluyveromyces marxianus* or *K. marxianus*) are well known for their ability to utilize a large variety of compounds as carbon and energy sources for growth. Since these strains are able to grow at high temperatures and exhibit high growth rates they are promising hosts for the industrial production of heterologous proteins.

Among the substrates that support growth are polysaccharides such as inulin, xylan and pectin, which are degraded by extracellular enzymes. Inulinase (EC.3.2.1.7) is an extracellular enzyme that enables the yeast to grow on fructans such as inulin and sucrose. The enzyme occurs in two forms, whereby part of the enzyme is secreted into the culture fluid as a dimer and part is retained in the cell wall as a tetramer. The relative amounts of the cell wall and supernatant enzyme depend on cultivation conditions, a situation similar to that of invertase (E.C.3.2.1.26) of *Saccharomyces cerevisiae*. The two enzymes differ in substrate specificity for inulin and sucrose, a

fact normally expressed in the S/I ratio (Vandamme *et al.*, 1983).

The pKD1 plasmid (Falcone *et al.*, 1986) originally found in *Kluyveromyces drosophilarum* (now regarded as a variety of *Kluyveromyces lactis*) belongs to the family of yeast double stranded circular plasmids, and does not confer any evident phenotype. Based on plasmid pKD1 several commercially attractive expression systems for high level expression of prochymosin (v/d Berg *et al.*, 1991) and human serum albumin (Fleer *et al.*, 1991) have been developed for the yeast *Kluyveromyces lactis*. As known from *S. cerevisiae*, a high copy plasmid based expression system has the advantage of supplying the host with a sufficient number of gene copies to obtain high-level expression, while integration into the genome in single or low copy number increases the mitotic stability under non selective growth conditions.

To combine the benefits of both expression systems, the concept of a multicopy integration system in the rDNA locus of *S. cerevisiae* has already been successfully proven (Verbakel, 1991). The potential of these constructs for stable, multicopy integration into the genome, has been demonstrated for different organisms, genes and auxotrophic markers (Lopes, 1990; Verbakel, 1991, Bergkamp *et al.*, 1991). Progress has been made elsewhere to stabilize a plasmid borne expression system (Fleer *et al.*, 1991), the adaptation of multicopy integration into the genome of *K. marxianus* is however a more favourable option; and is currently under investigation. Therefore the first *leu2* mutant of strain CBS 6556 has been made, in this specification named KMS1, in which the *LEU2* gene is inactivated through integration of a pPGK/Neomycin resistance (Neo^R) cassette (Bergkamp *et al.*, 1991).

Within the transferring process of the multicopy integration system from *S. cerevisiae* to *K. marxianus*, first a vector was developed, which is capable of integrating into the genome of *K. lactis* by targeted homologous recombination in the ribosomal DNA locus

(Bergkamp *et al.*, 1992). Using this vector system, the expression of an α-galactosidase gene from a fusion construct containing the *S. cerevisiae* GAL7 promoter, the SUC2 invertase signal sequence was obtained. With a maximum number of integrated plasmids of about 15, a level of about 250 mg/l α-galactosidase was obtained, with a secretion efficiency of about 95%. Compared to the ARS- or pKD1 derived *K. lactis* vectors containing the fusion construct, the multicopy integrants exhibit a considerably higher stability under non-selective growth conditions.

However, in addition to the importance of a stable, high copy, system the strengths and effectiveness of surrounding regulatory sequences seem to be crucial factors for high level expression, at least in *K. marxianus*. This is supported by results from the same group, where attempts to use the *S. cerevisiae GAL7* and *PGK* promoters for expression in *K. marxianus* have, in contrast to their effects in *K. lactis*, only led to an extremely low yield.

On the other hand, the homologous ORF1 promoter of killer plasmid k1 and the LAC4 promoter have already been successfully tested by different companies (Fleer *et al.*, 1991, v/d Berg *et al.*, 1991).

Yet another difficulty is the proficient secretion of recombinant protein by *Kluyveromyces*, especially when the protein is expressed in large quantities. Even though some heterologous secretion/signal sequences have been shown to be functional in *K. lactis*, as for example the human serum albumin prepro-sequence (Fleer et al., 1991), there is a strong demand for an efficient homologous signal sequence, especially from *K. marxianus*.

Description

Since inulinase is known to be expressed in very high concentrations under appropriate cultivation conditions in *K. marxianus*, the present invention is directed in particular at the cloning of regulatory regions, such as

the promoter sequence and the signal sequence of the inulinase gene as promising components for the development of an expression system.

This invention therefore relates generally to a nucleic acid sequence derivable from a yeast and comprising at least a regulatory region derivable from a gene encoding a polypeptide having inulinase activity. Said nucleic acid sequence can be combined with nucleic acid sequences encoding other homologous or heterologous genes to bring these genes under the control of at least one strong inulinase regulatory sequence.

"Nucleic acid sequence" as used herein refers to a polymeric form of nucleotides of any length, thus to both single and double stranded deoxyribonucleic acid (DNA) sequences, to ribonucleic acid (RNA) sequences, as well as to modifications thereof. In principle a single stranded nucleic acid DNA refers to the primary structure of the molecule.

In general the term "polypeptide" refers to a molecular chain of amino acids with a biological activity and does not refer to a specific length of the product and if required can be modified in vivo or in vitro. This modification can for example take the form of glycosylation, amidation, carboxylation or phosphorylation; thus, inter alia, peptides, oligopeptides and proteins are included. In this instance the polypeptide has inulinase activity.

Yet another major aspect of the present invention is related to the isolation, characterization and the use of the signal sequence of a polypeptide having inulinase activity, and parts thereof, for secretion of any overexpressed product from yeast, in particular from *Kluyveromyces*. A nucleic acid sequence according to the invention can therefore optionally further comprise a nucleic acid sequence encoding a secretory signal of inulinase.

The invention further relates to a vector containing the

nucleic acid sequences as described and also relates to micro-organisms containing said vectors or nucleic acid sequences.

The invention is also directed at modified sequences of the aforementioned nucleic acid sequences according to the invention, said modified sequences also having regulatory activity. The term "a modified sequence" covers nucleic acid sequences having the regulatory activity equivalent to or better than the nucleic acid sequence derivable from a yeast and comprising at least a regulatory region derivable from a gene encoding a polypeptide having inulinase activity. Such an equivalent nucleic acid sequence can have undergone substitution, deletion or insertion, or a combination of the aforementioned, of one or more nucleotides resulting in a modified nucleic acid sequence without concomitant loss of regulatory activity occurring. Such modified nucleic acid sequences fall within the scope of the present invention. In particular modified sequences capable of hybridizing with the non modified nucleic acid sequence and still maintaining at least the regulatory activity of the unmodified nucleic sequence fall within the scope of the invention.

The term "a part of" covers a nucleic acid sequence being a subsequence of the nucleic acid sequence derivable from a yeast and comprising at least a regulatory region derivable from a gene encoding a polypeptide having inulinase activity. The term "a part of" also covers a subsequence that is specific for the nucleic acid sequence derivable from a yeast and comprising at least a regulatory region derivable from a gene encoding a polypeptide having inulinase activity, said subsequence having a length of at least ten nucleotides and being capable of hybridizing to a regulatory region of a yeast inulinase gene under stringent conditions, said subsequence being suitable for use as a probe or a primer. The invention is in fact also directed at such use of a nucleic acid sequence according to the invention.

In particular the invention is directed at a nucleic acid sequence derivable from a yeast of the genus *Kluyveromyces*. A suitable example of a yeast from which a nucleic acid sequence according to the invention can be derived is a *Kluyveromyces* of the species *K. marxianus*. Of this species the strain *K. marxianus* var. *marxianus* is eminently suitable for deriving a nucleic acid sequence according to the invention. This strain, deposited in 1974 at the Centraal Bureau voor Schimmelcultures (CBS) in Baarn, The Netherlands under accession number CBS 6556, is freely available and is also known as NRRL 47571 and ATCC 26548. Preferably, the nucleic acid sequence according to the invention comprises at least a promoter as regulatory region. The nucleic acid sequence according to the invention can also comprise an enhancer sequence enabling a higher level of expression of any nucleic acid sequence operably linked to the promoter. The nucleic acid sequence can further either comprise an activating sequence upstream of the promoter (UAS) that can be activated by an inducer, said inducer being a compound that binds to said UAS, or a repressing sequence upstream of the promoter (URS) that can be derepressed by a component that inactivates the repressor being present at said URS. In older literature the term "operator" was sometimes used. Instead of this term UAS and URS are now often used. It is also possible for the nucleic acid sequence according to the invention to comprise a termination signal as regulatory region. Naturally, the nucleic acid sequence according to the invention can comprise one or more regulatory regions. A nucleic acid sequence according to the invention can comprise solely the promoter as regulatory region or a combination thereof with an enhancer, UAS or URS. A nucleic acid sequence according to the invention can also further comprise termination signal sequences, although these are not always required to end expression of the desired expression product. A nucleic acid sequence according to the invention can further comprise a sequence encoding a

secretory signal necessary for secreting a gene product from a yeast. This will be preferred when intracellular production of a desired expression product is not sufficient and extracellular production of the desired expression product is required. Secretory signals comprise the prepro or pre sequence of the inulinase gene for example. A secretory signal derivable from the inulinase gene of a *Kluyveromyces* yeast is particularly favoured. The specific embodiment of the nucleic acid sequence according to the invention will however depend on the goal that is to be achieved upon using a sequence according to the invention.

With the help of DNA oligonucleotides deduced from either existing protein sequencing data of inulinase from *K. marxianus* or newly obtained protein sequence analysis for example a 290 bp DNA fragment has been generated by use of the PCR technique, and the fragment has further been used for the isolation of chromosomal DNA fragments containing the whole inulinase gene of *K. marxianus* including the regulatory regions, such as the promoter, the signal sequence and the termination sequence. The invention is therefore in particular directed at a nucleic acid sequence derivable from a yeast and comprising at least a regulatory region derivable from a gene encoding a polypeptide having inulinase activity in any of the embodiments described above, said nucleic acid sequence comprising at least a part of the nucleic acid sequence of figure 5 or an equivalent nucleic acid sequence. The term "equivalent nucleic acid sequence" has the same meaning as given above for "a modified nucleic acid sequence".

Yet another aspect of this invention relates to the isolated nucleic acid fragment of *K. marxianus* containing an open reading frame encoding 556 amino acids, with nucleotides encoding a prepro-peptide sequence of 23 amino acids at the amino terminus. The calculated molecular weight of the corresponding gene product is 62.5 kDa, which is in good agreement with the 64 kDa experimentally

determined for the corresponding polypeptide (Rouwenhorst *et al.*, 1990).

A further aspect of the invention is directed at processes for the construction of either episomal or integrating expression vectors containing the described regulatory sequence or sequences. In the given examples the expression and secretion potential of the obtained INU promoter and the INU signal sequences have been tested by constructing a variety of new vectors for expression of a heterologous α-galactosidase gene in *Kluyveromyces*. The resulting constructs were tested in *K. marxianus*, variety *marxianus*.

Yet another aspect of this invention relates to a method of transforming a *Kluyveromyces* strain capable of producing a heterologous protein through fusion with the prepro- and the pre-part of the homologous inulinase signal sequence. High expression levels and nearly complete secretion were obtained with all episomal plasmids that were constructed. A strain, transformed with a construct containing the whole prepro-sequence secreted up to 150 mg/L enzyme when grown in shake flask, which is an approximately 100 fold increase compared to the vectors containing non homologous *S. cerevisiae* promoters and signal sequences.

In another embodiment of the present invention the PCR technique is used in combination with the use of class IIS restriction enzymes, to facilitate primarily the functional new recombination of the described DNA fragments.

As a typical example, *Bsp*MI, a constituent of the group of IIS restriction endonucleases, cuts every DNA sequence 4 bp 5' of the specific recognition site "ACCTGC", thereby generating 5' N4 protruding ends (reviewed by Szybalski et al, 1991). The advantage of these enzymes, particularly in combination with PCR, is the nearly complete independence from a given sequence within modern molecular working procedures. Introduction of the recognition sequence into the non priming part of a primer used for PCR, allows subsequent generation of any desired end of the PCR fragment.

Furthermore, this invention relates to a process for producing a desired expression product wherein a host cell comprising a vector according to the invention is cultured under conditions enabling the expression of the structural gene and optionally the isolation of the desired expression product.

The invention also relates to a method for producing RNA, wherein a host cell comprising a recombinant nucleic acid sequence according to the invention is cultured under conditions enabling production of the RNA, whereby said recombinant nucleic acid sequence further comprises a regulatory region operably linked to DNA encoding a specific RNA sequence not encoding a specific protein. Such a process can for example be used to produce RNA itself as the desired expression product, or to produce RNA that influences the formation of at least one metabolite as the desired expression product. The amount of metabolite can be increased by using the regulatory region or regions according to the invention (as described above in various suitable embodiments) in combination with a nucleic acid sequence encoding a protein that influences the formation of the metabolite.

The production of anti-sense RNA, that binds to sense RNA encoding an expression product, can be used for decreasing the amount of said desired expression product, whereby the latter can be a specific protein or a protein influencing the formation of a metabolite.

A nucleic acid sequence according to the invention can therefore also comprise an anti-sense nucleic acid sequence in combination with one or more of the regulatory regions according to the invention.

A process for producing RNA according to the invention can be directed at the production of an RNA sequence that functions as a flavouring component. The nucleic acid sequence according to the invention is therefore not only to be considered useful for overexpression of a proteinaceous gene product but also for producing RNA.

Brief Description of the Figures

Figure 1.

Protein sequence analysis of two forms of inulinase from *K. marxianus* after CNBr-digestion. Fragment 1 in figure 1 corresponds to Seq. ID. No. 1 and fragment 2 in figure 1 corresponds to Seq. ID. No. 2. The expected cleavage after a Met- residue is indicated by an arrow, small letters in the sequence indicate very likely residues. Identification is based on either homology with invertase (n) or strong suspicion.

Figure 2.

a) DNA oligonucleotides derived from the amino acid sequence of the internal CNBr-fragments 1 and 2. Fragment 1 in figure 2a corresponds to Seq. ID. No. 3 and fragment 2 in figure 2a coreesponds to Seq. ID. No. 4.

b) DNA probes from the mature N-terminus of secreted inulinase. The number of nucleotides is given in brackets, as well as the abbreviations used in the text (probe KLM 04 corresponds to Seq. ID. No. 5, probe KLM 05 corresponds to Seq. ID. No. 6, probe KLM 08 corresponds to Seq. ID. No. 7 and probe KLM 09 corresponds to Seq. ID. No. 8). In cases where mixed oligonucleotides are used during DNA synthesis, the corresponding letters are given; the orientation of the DNA oligonucleotides is mentioned.

Figure 3.

Nucleotide sequence (Seq. ID. No. 15) of the 280 nucleotides long PCR fragment of the N-terminal coding region of the inulinase gene in pTZ18R. The localization of two corresponding PCR primers; as well as their code, is given. In the line "seq" the experimentally determined amino acid sequence of inulinase from *K. marxianus* is given, The deduced amino acid sequence (Seq. ID. No 16) is mentioned in the line below the DNA sequence, here, all amino acids identical to amino acids of *Saccharomyces*

cerevisiae invertase are underlined.

Figure 4.

First restriction endonuclease cleavage map of the region around the inulinase gene of *K. marxianus*. Restriction sites were located through *Kpn*I double digestions. Not all restriction sites of the given restriction enzymes are mentioned. 0 kbp mark refers to the 5' end of the coding sequence of inulinase. In the upper part of the figure about 22 Kbp are mapped, while the lower part displays a refinement of approximately 2.5 Kbp around the target sequence.

Figure 5.

Nucleotide sequence (Seq. ID. No. 9) of the inulinase gene (*INU1*) of *Kluyveromyces marxianus*. The TATAAA box, transcription start sites, the putative MIG1 binding site as well as the predicted recognition site for the signal peptidase (G-V-S-A-↑-S-V-I) and the processing site for a KEX2-like endoprotease (K-R-↑-) are indicated. Numbering starts with the ATG start codon, the deduced amino acid sequence (Seq. ID. No. 10) is given in one letter code below the coding part.

Figure 6.

Autoradiogram of the primer extension assay. Results are shown for the primer extension assays in the presence of [α-^{32}P]dCTP with total RNA from repressed [1] and derepressed grown cells [2]. The size of the obtained fragments is indicated.

A: assay with primer p21T;

B: assay with primer p16T;

C: specificity control; both primers with total RNA from *S. cerevisiae*.

For further details see text.

Figure 7.

Schematic representation of the construction for the inulinase promoter/signal sequence link to α-galactosidase with the help of PCR generated fragments. The beginning of mature protein and the first amino acid of the pre-protein is indicated by arrows. Digestion of plasmids such as for example pSK1 with *Eco*RI and *Eag*I removes the DNA part comprising the GAL7 promoter and the SUC2 signal sequence in such a way, that an in frame fusion of inulinase signal sequences with the α-galactosidase gene was directly possible. For the in frame fusion of the whole prepro-sequence to α-galactosidase an oligonucleotide complementary to the coding strand was used as PCR primer, said oligonucleotide comprising the recognition site of *Bsp*MI. After PCR, digestion of the product with *Eco*RI and *Bsp*MI created sticky ends, that were compatible with the ends of the original vector, for example pSK1. By changing the hybridizing part of the PCR primer a similar fragment was obtained for the in frame connection of the pre-sequence to α-galactosidase.

Figure 8.
Schematic representation of the construction routes of plasmids, suitable for the expression of a heterologous gene, here for example α-galactosidase, within *K. lactis*. The construction route is only given for an episomal plasmid based expression system. The sequence of the α-galactosidase gene is indicated in black, shadowed areas indicate yeast sequences, solid lines are bacterial sequences, the direction of transcription is indicated.

Figure 9.
Schematic representation of the construction route of plasmids pUR2431 and pUR2432, examples of construction intermediates for the expression of α-galactosidase from the INU promoter, containing either the intact prepro-signal sequence or only the pre- part of the inulinase signal sequence. The promoter sequence is shaded, the empty

box indicates both versions of the signal sequence, the example of a heterologous gene, here α-galactosidase is given as a black box. The direction of transcription is indicated by arrows.

Figure 10.
Schematic representation of the construction of plasmids pUR2433, pUR2434, pUR2435 and pUR2436, all variants of episomal expression plasmids for the expression of α-galactosidase in *Kluyveromyces* strains, having *leu2* determined auxotrophy.

Figure 11.
Schematic representation of the construction of pUR2437 and pUR2438, vectors for integrating multiple copies of a homologous or heterologous gene, such as α-galactosidase, into the rDNA of *K. marxianus*. The overall structure of one rDNA unit as well as the 3.5 kbp *Eco*RI fragment actually used are drawn schematically.

Figure 12.
Sequence of the *K. marxianus URA3* gene (corresponding to Seq. ID. No. 11) and its deduced amino acid sequence (corresponding to Seq. ID. No. 12).

Figure 13.
The construction of plasmid pKMU2, which was used for the construction of a food-grade K. marxianus *leu2* mutant. (A) Plasmid pKML1 contains a K.marxianus *LEU2* gene on a 5 kb *Eco*RI fragment. (B) . Plasmid pKMU2, where the intact *LEU2* gene is replaced by a *leu2::URA3* disruption. The small boxes indicate the *URA3* promoter and terminator regions.

Figure 14.
Structure of plasmid pKUR2431 (A) and the chromosomal organization of the *INU1* locus after integration of the plasmid at the *Xho*I site (B).

Figure 15.
Structure of plasmids pUR2439 and pUR2440 which contain *K. marxianus* DNA 5' to the previously cloned sequences. These plasmids are based on pBluescript (Stratagene) and the inserted *K. marxianus* DNA is shown as the dark shaded boxes.

Figure 16.
The sequence of the 479 bp *Eco*RI fragment 5' to the previously cloned sequences. The sequence (corresponding to Seq. ID. No. 13) begins with the 5' *Eco*RI site from the previously cloned *INU1* DNA.

Figure 17.
Sequence (corresponding to Seq. ID. No. 14) of primer INUT used for sequencing across the *Eco*RI site 5' to the inulinase gene.

Figure 18.
Structure of plasmid pUR2445 which contains the 470 bp *Eco*RI fragment from pUR2440 5' to the *INU1* sequences in pUR2434. The location and orientation of the approximatley 470 bp *Eco*RI fragment is indicated.

Experimental

The following experimental section is offered by way of example and should not be considered a limitation of the scope of the invention.

Molecular biological procedures

Mostly standard methods were used as described in Sambrook, J., Fritsch, E.F., & Maniatis, T., 1989. Molecular Cloning. A laboratory Manual. Second edition. Cold Spring Harbour Laboratory Press. Any modifications used are described below.

Strains, plasmids and growth conditions

E. coli strain JM109 (endA1, recA1, syrA96, thi, hsdR17, rk^-, mk^+, relA1, supE44, Yanish-Perron *et al.*, 1985) was used for amplification of plasmids. Transformation of JM109 was carried out according to Cohen *et al.*, 1973.

K. marxianus var. *marxianus* CBS 6556 (= ATCC 26548) was obtained from the Yeast Division of the Centraalbureau voor Schimmelcultures, Delft, The Netherlands, and maintained on YEPD agar (1% yeast extract, 2% peptone, 2% glucose, 2% agar) slopes.

Genomic DNA was isolated from a 200 ml YEPD overnight culture and incubated with lyticase (Sigma Chemical Company) from *Arthrobacter luteus* according to the manufacturer.

Total DNA was isolated as described by Struhl *et al.* (1979).

The *K. marxianus* strains were transformed with the plasmids pUR2431 up to and including pUR2445. Transformation of the *Kluyveromyces* strains was performed as described by Carter *et al.*, 1988. Transformants were recovered on selective YNB-plates (0.67% YNB, 2% glucose, 2% agar) supplemented with the essential amino acids (tryptophan 20µg/ml or leucine 20µg/ml). The same liquid medium was used for precultures, cultivated twice overnight at 30°C and diluted 1:10 in YPmedium containing 5% sucrose (YPS) for derepression of the INU promoter.

Example 1. Generation of DNA oligonucleotides

To acquire a set of DNA oligonucleotides for PCR a set of mixed DNA oligonucleotides corresponding to the recently determined N-terminal amino acid sequence of forms I and II was synthesized [Rouwenhorst *et al.*, 1990].

As a potential source for further DNA-probes and to apply the PCR technique, the amino acid sequences of two internal CNBr fragments from the secreted inulinase form I and the cell wall bound inulinase, form II were determined. Therefore the reduced and carboxyamidomethylated proteins

were subjected to overnight incubation in 70% formic acid in the dark under N_2. After addition of water the mixtures were freeze dried and yielded the CNBr-digests of inulinase forms I and II respectively.

Separation of the obtained fragments was achieved by reversed phase chromatography using a Bakerbond C4 wide pore column (4.6*250mm) mounted in a Waters HPLC. Elution was achieved using a linear gradient of acetonitrile in 0.1% Trifluoroacetic acid in water. Detection was carried out at 214 and 254 nm.

The chromatograms obtained from digests of inulinase forms I and II showed the presence of a number of poorly resolved peaks; the overall pattern however was similar for both digests and enabled collection of two fractions from both runs which were subjected to sequence analysis. The outcomes of the runs are given in Fig.1 (corresponding to Seq. ID. No. 1 and 2).

No differences could be detected in the amino acid sequences derived for the isolated fractions between forms I and II. Peptide bond hydrolysis has occurred in one case at a C-terminal Trp residue which is rare but not impossible [fragment 2 in Fig.1/Seq. ID. No. 2] under the given circumstances.

The nucleotide sequence was selected in such a way, that PCR could generate the genetic information of the intervening sequence. From these sequence results sets of mixed oligonucleotides were synthesized using an Applied Biosystems 380 A synthesizer. The sequence of these DNA oligonucleotides is given in Fig. 2 (corresponding to Seq. ID. No. 3 to 8).

Example 2. Cloning the 5' coding region of inulinase

With two of the obtained DNA oligonucleotide probes from the N-terminal and internal protein sequence [KLM09, resp. KLM06] PCR amplification on total *K. marxianus* genomic DNA was carried out (Perkin Elmer Cetus DNA Thermal Cycler) in 100 μl 10 mM Tris HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$,

0.001% gelatine, with 0.2 mM of each dNTP, 100 pmol of the DNA oligonucleotides KLM06 and KLM09, approximately 0.5 μg of *Bam*HI digested DNA and 1U of Amplitaq polymerase. Incubation parameters were set as follows: 32 cycles/ 1 min 95 °C/ 2 min 50 °C/ 2.5 min 72 °C.
The reaction formed a specific 290-bp fragment, which was subcloned into the *Sma*I site of the *E. coli* plasmid pTZ18R (Mead et al, 1986) and introduced into *E. coli* JM109 by the transformation protocol described by Chung *et al.* (1989). One of the positive clones, designated pUR2415, was further characterized, the DNA isolated and purified according to the Qiagen (Qiagen Inc., Chatsworth California) protocol and subsequently characterized by DNA sequencing. The sequence of the PCR clone is given in Fig. 3 (corresponding to Seq. ID. No. 15).

Comparison of the gained DNA sequence with known sequence data of invertase from *S. cerevisiae* further confirmed the authentic origin of the PCR product, since it displayed a significant sectional homology with the supposed N-terminus of invertase as described by Rouwenhorst *et al.* (1990).

Example 3. Restriction map of the DNA around the inulinase gene

The *Nco*I-*Bam*HI fragment of pUR2415 was further used as a ^{32}P labelled probe for the construction of a physical map of the DNA region around the 5' coding sequence of the inulinase gene. Therefore chromosomal DNA of *K. marxianus* was digested with several restriction enzymes separately and in combination. After electrophoresis of DNA fragments the gel was placed for 15 min in 0.25 M HCl, 15 min in 0.4 M NaOH, 0.6 M NaCl and 15 min in 0.5 M Tris, 1.5 M NaCl. The DNA was transferred onto Hybond N-filters (Amersham International plc.) by vacuum blotting (LKB 2016 Vacugene) for 2 hours in 10xSSC (1.5M NaCl, 0.15M Na_3citrate) and finally UV crosslinked for 15 min.

By using *Kpn*I double digestions the resulting signals were arranged into a first physical map spanning about 25 Kbp, (given in Fig.4).

Example 4. Cloning of the inulinase gene

Results of the chromosomal restriction analysis revealed two positive overlapping DNA fragments of 2.0 Kbp for *Eco*RI and 4.0 Kbp for *Pst*I, respectively. To isolate clones containing the inulinase promoter, the signal sequence and the polyadenylation/termination sequences both digested DNA pools were subcloned into pTZ19. Therefore about 8 µg of chromosomal DNA was digested with *Eco*RI and *Pst*I separately and resolved by agarose gel electrophoresis. DNA fragments of about 2.0 Kbp from the *Eco*RI digest and the fragments between 3.5 and 4.5 Kbp from the *Pst*I digest were isolated from the gel and purified with the Geneclean II kit (Bio 101 Inc). A small amount of both digestions was again loaded onto an agarose gel, the bands transferred to Hybond membrane, and hybridized with the ^{32}P labelled PCR fragment to verify the presence of the hybridizing band within the isolated pool. Since both fractions contained the corresponding DNA fragments the isolated *Eco*RI and *Pst*I DNA fragment pools were ligated into the resp. digested pTZ19 plasmids and transformed into *E. coli* JM109 by standard procedures. The colonies obtained were subjected to colony hybridization after replica plating them onto Hybond-N filter (Amersham International plc) and plasmid amplification on LB-plates with containing 500 µg chloramphenicol ml^{-1}. After 8 hours incubation at 37°C each filter was subjected to the following wash procedure: 5 min.in 1.5 M NaCl, 5 min in 0.5M NaOH and twice in 1.5M NaCl, 0.5M Tris HCl. The DNA was finally fixed to the filter by UV crosslinking.

Hybridisation was done in 50mM Tris pH 7.4, 10 mM EDTA pH 7.0, 1M NaCl, 0.5% SDS, 0.1% Na-pyrophosphate, 0.2% ficoll, 0.2% polyvinylpyrolidone, 0.2% BSA and 0.01 mg denaturated salmon sperm DNA at 68 °C. The added [α-^{32}P]dCTP (Amersham

International plc; 370 MBq/mL; 110 TBq/mmol) labelled DNA probe was prepared by using a Multiprime DNA labelling kit from Amersham Corporation, purified by elution over a Sephadex G-50 column in TES (10mM TriHCl pH 7.4, 1 mM EDTA pH 8.0, 0.2% SDS) and then denatured by incubation for 2 minutes at 100 °C prior to use. After overnight incubation the filters were washed 2x for 20 min in 2xSSC, 0.1% SDS, 0.1% Na-pyrophosphate; 2x 20 min in 0.1x SSC, 0.1% SDS, 0.1% Na-pyrophosphate at 68°C. Positive clones were detected after overnight exposure of the dried filters with a Kodak-X-ray film.

To verify the specificity of the obtained spots, filters enclosing presumably positive clones were washed for 30 min in 2% SDS at 90°C and rehybridized with a DIG labelled DNA probe of the PCR fragment, by using the DIG luminescent detection kit from Boehringer Mannheim according to the manufacturer's protocol.

Plasmid DNA was isolated from putative positive colonies, digested with appropriate enzymes and analysed by Southern hybridization as follows: after electrophoresis of the digested DNA the gel was placed for 15 min in 0.25 M HCl, 15 min in 0.4 M NaOH, 0.6 M NaCl and 15 min in 0.5 M Tris, 1.5 M NaCl. The DNA was transferred onto Hybond N-filters (Amersham International plc.) by vacuum blotting (LKB 2016 Vacugene) for 2 hours in 10xSSC and finally UV crosslinked for 15 min. In this experiment the probe was labelled through random primed incorporation of DIG-UTP according to the protocol of the manufacturer (DIG DNA labelling kit, Boehringer Mannheim). After overnight hybridisation with the denaturated probe at 68 °C in 5x SSC, 0,1% N-laurylsarcosine, 0,02% SDS, 1% blocking agent) filters were washed as described in the manufacturers instructions.

The fragments found in the plasmids which hybridized with the probe were in reliable agreement with the physical map given in Fig. 4. One of the 2.0 Kbp *Eco*RI insert containing pTZ19R plasmids, designated pUR2421, and one 4.0 Kbp *Pst*I fragment comprising pTZ19R plasmid, designated pUR2422,

were further utilized to determine the DNA sequence of the total inulinase gene.

Example 6. DNA sequence analysis of the complete inulinase gene

DNA sequencing was mainly done as described by Sanger *et al.*, 1977, and Hsiao *et al.*, 1991, using the Sequenase version 2.0 kit from United States Biochemical Company, according to the protocol with T7 DNA polymerase (Amersham International plc) and [α-^{35}S]dATP (Amersham International plc: 370 MBq/ml; 22 TBq/mmol).

The complete sequence was determined from the recombinant plasmid pUR2421 and from the plasmid pUR2422 by subcloning fragments and by primer walking strategy. Both DNA fragments showed the expected overlap. In summary, 3223 bp were sequenced on both DNA strands, including the promoter region extending over about 0.75 Kbp upstream of the putative start codon and a sequence of about 0.83 Kbp behind the putative stop codon, including the putative polyadenylation site and termination regions. The sequence is given in Fig. 5 (corresponding to Seq. ID. No. 9 and 10). Sequence comparison of the coding part of the inulinase gene of the present invention showed about 98% homology with the very recently published inulinase coding sequence of *K. marxianus*, ATCC 12424 (Laloux *et al.*, 1991). The homology is less striking for the 50 nucleotides before the prepro-sequence given in the same publication, corresponding to about one third of the leader sequence before the start codon from pUR2421. This variation is probably due to strain variations.

At the amino acid sequence level, invertase and inulinase display a homology of 69%, a homology, which is even higher than the homology between invertases from *S. cerevisiae* and *Schwanniomyces occidentalis*. Therefore, both enzymes should be treated as variations of the same enzymatic activity rather than as different enzymes.

Since the N-terminus of secreted mature inulinase was

identified by protein sequencing, it was easy to distinguish the coding part of the precursor protein, having a deduced 23 amino acid sequence displaying some characteristic prepro- features. 270 bp in front of the supposed ATG start codon a TATAAA box was identified, indicating the presence of a promoter element.

Example 7. Determination of transcription start

To test the functionality of the detected promoter structure and to identify the transcription start points of the cloned inulinase fragment, primer extension experiments were carried out. For the primer extension assay two DNA oligonucleotides were used, complementary to nucleotides -98 to -84 and 18 to 48 of the given DNA sequence given in Fig.5 (which corresponds to Seq. ID. No. 9).

primer p21T 5'- AGC ACT GAC TCC TGC CAA TGG AAG CAA GAG (Seq. ID. No. 17)

primer p16T 5'- TCT CTA TGG CAT AGA GA (Seq. ID. No. 18)

To further confirm the specificity of the signals, two different total RNA preparations for the reverse transcription were chosen. Therefore, *K. marxianus* cells were grown as described in YPS medium (Rouwenhorst at al., 1990) under non-repressive conditions. From a 100 ml culture, cells were harvested and total RNA was isolated as described (Koehrer *et al.*,1991). A reverse transcriptase reaction was carried out in the presence of either [α-^{32}P]dCTP or [α-^{35}S]dATP. For each experiment about 10 μg RNA and 100pg of primer were dissolved in 40 μL hybridisation buffer, containing 50 mM Tris.HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT and 40 U RNA'se inhibitor (Boehringer Mannheim). The mixture was incubated at 65 ^{o}C for 5 min. and slowly cooled down to room temperature. 1 μL [α-^{32}P]dCTP (Amersham International plc: 370 MBq/ml; 110 TBq/mmol) or [α-^{35}S]dATP,1 μL (25 U) reverse transcriptase (Biolabs), dATP, dGTP and dTTP to a final concentration of 0.1 mM, dCTP to a final concentration of 0.01 mM were added

to a final volume of 50 µl and incubated for 1 hour at 37 °C (for reactions containing [α-^{35}S]dATP the nucleotide concentrations were 0.1 mM for dGTP, dTTP and dCTP and 0.01 mM for dATP). The mixture was precipitated with ethanol and subsequently loaded onto a 5% Polyacrylamide gel together with the DNA sequence reactions generated from the same primers (Fig.6).

In all experiments, three dominant signals emerged, coinciding in each case to T_{-174}, C_{-170} and C_{-167} of Fig.5 (which corresponds with Seq. ID. No. 9). These nucleotides are located about 100 nucleotides behind the **TATAAA** box [position -276 to -271 in Fig.5]. The results of this part of the invention therefore relate the start of transcription for the inulinase gene within the region TAATCAGCAATT, defining the length of the uncommonly long 5' non-coding sequence as 174, 170 and 167 nucleotides. Directly behind this transcription initiation region we recognized a sequence [TAAATCCGGG, nucleotides -163 to -153 in Fig.5 (which corresponds with Seq. ID. No. 9)] that perfectly matches the MIG1 binding consensus sequence of *S. cerevisiae SUC2, GAL4* and *GAL1* genes [WWWWTSYGGGG] (Nehlin *et al.*, 1991). The *MIG1* gene product is known to be involved in glucose repression of the *GAL* genes and directly controls *SUC2* expression in *S. cerevisiae* (Nehlin *et al.*, 1990). In contrast to the presumably related *SUC2* gene, where the MIG1 binding site is located at -446 to -435, this sequence motif is closer to the start codon in *K. marxianus* and more similar to the location of the MIG1 site in the *GAL1* and *GAL4* promoters. But in contrast to the regulation of the GAL gene family of *S. cerevisiae*, inulinase and invertase promoters seem to be solely regulated by glucose repression; a fact, allowing the construction of a strong, non-repressible promoter by exchanging the putative MIG1 DNA binding site in the inulinase promoter.

The operative importance of the sequence around the AUG start codon during initiation of translation in eukaryotes

is still a point of discussion, and has led to the formulation of a consensus sequence for *S. cerevisiae* (Hamilton *et al.*, 1987). The inulinase gene shows little homology with this sequence and even the formulation of a preliminary *Kluyveromyces* ATG context consensus sequence does not improve the homology significantly. Taking the known high expression level of inulinase in the natural host into account, the idea of improving protein expression through adaptation of the AUG context in *Kluyveromyces* mRNA on the basis of the present information seems to be less striking.

Example 8. PCR of *Kluyveromyces* regulatory sequences

Sequence analysis of the prepro-sequence of the leader region of the cloned inulinase promoter in pUR2421 revealed 5 amino acids after a predicted recognition site for the signal peptidase (G-V-S-A-↓-S-V-I) (Van Heijne, 1986) also a putative processing site for a KEX2 like endoprotease (..K-R-↓-..) (Fuller *et al.*, 1988). To test the functional importance of the prepro-sequence, DNA oligonucleotides were synthesized, one creating the complete prepro-inulinase sequence, and the second appropriate for the direct, in frame, attachment of the DNA coding for the inulinase signal peptide to a given coding sequence. Two oligonucleotides containing the *Bsp*MI recognition site, one with sequence information for new restriction sites and the complementary sequence for the DNA from the putative KEX2 protease site, and the second complementary for the signal peptidase cleavage site, were utilized for the generation of suitable promoter/signal sequence fragments by PCR. For the assembly of promoter and secretion signal fragments from the inulinase gene, PCR amplification of this part of the inulinase gene was performed. Thereby, the primers were conceived in such a way that perfect couplings with mature alpha-galactosidase were obtained. Two versions were made, one with the inulinase pre-pro signal sequence and the other with only the inulinase putative pre signal

sequence (see fig.7.). The following primers were used:

INP 01: 5'-GGAATTCTCAAACCGAAATG-3' (Seq. ID. No. 19)
INP 02: 5'-CCCAAGCTTACCTGCCATGGGCCCTCTTGTAATTGATAACTG-3' (Seq. ID. No. 20)
INP 03: 5'-CCCAAGCTTACCTGCATGCGGCCGCACTGACTCCTGCCAATG-3' (Seq. ID. No. 21)

Two PCR-mixes of 100 µl were made, each containing 40 pg pUR2421 cut with *Kpn*I, 1.5 mM MgCl2, 1U AmpliTaq polymerase and buffers and NTP's as appropriate. One reaction mixture contained 100 pmoles of INP 01 and INP 02, the other 100 pmoles INP 01 and INP 03. 25 cycles were performed in a Perkin Elmer Cetus DNA Thermal Cycler, each cycle 1:00 min 95°C, 1:30 min 48°C, 2:00 min 72°C. Afterwards, they were treated with proteinase K (Crowe *et al.*, 1991) before digestion with *Eco*RI and *Hind*III was performed (48 hours at 37°C). The fragments were then isolated from a gel, and ligated with pTZ19R digested with *Eco*RI and *Hind*III. The resulting plasmids, pUR2427 and pUR2428 (INP 01/INP 02 and INP 01/INP 03 PCR-products respectively), were transformed into *E. coli* and several colonies for both constructions were cultivated, the plasmids isolated, purified and the sequence confirmed by DNA sequence analysis.

Example 9. Construction of *K. lactis* expression plasmids

The *E. coli* - *Kluyveromyces* shuttle vector pSK1 is a pKD1 derivative (Chen *et al.*, 1989, Bianchi *et al.*, 1987) and contains a unique *Eag*I site at the junction between the *S. cerevisiae SUC2* (invertase) signal sequence and the α-galactosidase gene. The vector also comprises a *K. lactis TRP1* gene as selectable marker in *K. lactis* and the ampicillin resistance gene as selectable marker in *E. coli*.

Digestion of said plasmids pUR2427 and pUR2428 with *Eco*RI and *Bsp*MI produces two fragments, which could be easily

ligated with the *Eco*RI/*Eag*I digested *Kluyveromyces* vector pSK1, thereby only replacing the *Eco*RI-*Eag*I fragment, containing the *GAL7* promoter and *SUC2* signal sequence with the *Eco*RI/*Bsp*MI fragments from pUR2427 and pUR2428. This results in two vectors, one with the DNA sequence encoding the expected signal peptide (= pre-sequence) directly linked to the α-galactosidase gene (pUR2429), and a second with the DNA encoding the natural prepro-sequence in frame linked to the α-galactosidase gene (pUR2430).

Said episomal plasmids could immediately be used to transform the trp$^-$ mutant strain of *K. lactis*, for example by electroporation (Bolen *et al.*, 1990) in which strain pSK1 derivatives are known to be stably maintained. Expression and secretion of α-galactosidase could be determined under induced and non induced conditions by known procedures (Verbakel, J., 1991).

By digestion of said plasmids with *Eco*RI and *Hin*dIII a fragment comprising both the promoter and the DNA encoding the leader sequence, as well as the α-galactosidase gene can be transferred into existing vectors for targeted homologous recombination into the *K. lactis* rDNA (Bergkamp *et al.*, 1992). The potential of these constructs for stable, multicopy integration into the genome, has been shown for different organisms, genes and auxotrophic markers (Lopes, 1990; Verbakel, 1991, Giuseppin *et al.*, 1991). Substitution, for example, of the GAL7 promoter and *SUC2* signal sequence in the plasmids pMIRKGAL-T▲1,2 and 3 for said inulinase promoter prepro- or pro-sequences, followed by transformation of the constructs into *K. lactis* MSK110 (a, uraA, trp1::*URA3*), should give high and stable expression and secretion of α-galactosidase.

DNA fragments comprising the whole given promoter sequence or functional parts thereof, but without the prepro-sequence coding part can also be used for intracellular overexpression of homologous and/or heterologous genes.

Example 10. Construction of *K. marxianus* episomal plasmids

To evaluate the function of the obtained DNA sequences in the natural host, the pTZ19R derivatives pUR2427 and pUR2428 were digested with *Eco*RI and *Bsp*MI, to release the PCR fragments, which were further used to simply replace the *Eco*RI/*Eag*I fragment in pSY9, an *E. coli/ S. cerevisiae* shuttle containing a unique *Eag*I site at the junction between the *S. cerevisiae SUC2* (invertase) signal sequence and the α-galactosidase gene and an *Eco*RI site in front of the GAL7 promoter; the vector also comprises a *LEU2* gene copy from *S. cerevisiae* and the ampicillin resistance and the MB1 origin for maintenance and selection in *E. coli* (M. Harmsen, unpublished). The two variant vectors, one with the direct connection of the expected signal peptide to the α-galactosidase gene (pUR2432), and the second with the natural prepro-sequence in frame linked to the α-galactosidase gene (pUR2431) are not able to replicate in *K. marxianus* (Figure 9). Finally, for obtaining the *Kluyveromyces* episomal expression vectors, the naturally occurring plasmid pKD1 was linearized with *Eco*RI and ligated into the *Eco*RI site of pUR2331 and pUR2332, thereby yielding 4 new plasmids, with the DNA encoding either the pre- or the prepro-sequence and in each case both possible orientations of the pKD1 vector backbone (pUR2433 - pUR2436).

Example 11. Construction of *K. marxianus* integrating plasmids

To obtain mitotically stable integration into the ribosomal DNA locus of the *K. marxianus* genome homologous rDNA sequences were used to target the integrating linearized vector into the rDNA locus. In addition to its use in *S. cerevisiae,* this approach has been successfully proven for multicopy integration into the rDNA locus of *K. lactis* using the vectors pMIRKM1 and pMIRKM2 (Bergkamp *et al.*, 1992). These plasmids include either the *LEU2* or the *LEU2d*

genes of *S. cerevisiae* and homologous rDNA sequences from *K. marxianus*. Since the heterologous *LEU2d* gene apparently is not able to functionally complement the *LEU2* gene disruption in the *K. lactis* strain (Bergkamp *et al.*, personal communication) only the vector with the intact LEU gene was used for further *K. marxianus* constructions. A very recently cloned 3.5 kb *Eco*RI fragment of the *K. marxianus* rDNA, containing the 3' end of the 17S rDNA-, the 5.8S rDNA-and the 5' end of the 26S rDNA gene, which has recently been cloned, but has not yet been completely sequenced (Bergkamp, personal communication) was ligated into the *Eco*RI sites of pUR2431 and pUR2432. After transformation into *E. coli* the transformants were found to contain only plasmids in which the inulinase- α-gal expression cassette was joined to the 26S rDNA gene part (pUR2437 comprising the prepro-sequence, and pUR2438 containing the pre-sequence).

Example 12. Expression and secretion in *K. marxianus*

The outcome of the complete construction process was 6 different expression plasmids for *K. marxianus*, with the following characteristics:

pUR2433: episomal vector;
INU promoter + prepro-sequence + α gal in orientation I

pUR2434: episomal vector;
INU promoter + prepro-sequence + α gal in orientation II.

pUR2435: episomal vector;
INU promoter + pre-sequence + α gal in orientation I

pUR2436: episomal vector;
INU promoter + pre-sequence + α gal in

orientation II.
pUR2437: integration vector;
INU promoter + prepro-sequence + α gal in orientation I.
pUR2438: integration vector;
INU promoter + pre-sequence + α gal in orientation I.

The four episomal plasmids and the two integration vectors were (after linearization at the unique *Xba*I site) transformed to the existing *K. marxianus leu2* strain KMS1 by known procedures (Carter at al.; 1988). In this strain, the gene coding for ß-isopropylmalate dehydrogenase was inactivated through integration of a dominant selection marker [G_{418}] under the control of the PGK promoter. The resulting strain is leu^-, Neo^R. Some of the acquired clones were grown overnight at 37°C in YNB medium and diluted 1:10 into 50 ml of YPS in a shake flask and grown for 48 hr at 37°C under non repressing conditions. The expression level of α-galactosidase was determined enzymatically intra- and extracellularly by known procedures (Verbakel, 1991; Giuseppin *et al.*, 1991). The copy number was preliminary estimated by Southern blot analysis.
The α-galactosidase expression assays confirmed the benefit of homologous regulatory sequences for high level expression in *K. marxianus*. Application of the inulinase DNA promoter sequence increased the expression of α-galactosidase up to 150 fold, compared to experiments, in which the *S. cerevisiae* PGK-or GAL7 promoters were used (data not shown). Moreover, the natural connection of this promoter to the corresponding signal sequence, led to nearly 100% secretion of the heterologous protein. Here, the use of the complete precursor sequence, including the S-V-I-N-Y-K-R pro-peptide sequence appears not only to increase the amount of secreted α-galactosidase, but also the amount of protein produced. This finding is in some conflict with the conclusion given by Fleer *et al.*, 1991,

where the deletion of the pro sequence of human serum albumin (HSA) did not influence the capacity of *Kluyveromyces lactis* to express and secrete rHSA. On the other hand, both experiments manifest, that the final proteolytic removal of the pro peptide from the mature product, presumably by a KEX2 equivalent protease, is not a rate limiting step. Whether the pro-sequence plays an appreciable role in secretion, translation or mRNA stabilization, cannot yet be decided. Some influence of the orientation of the DNA cassette within the plasmid has been found, at least for the episomal expression systems (compare pUR2433 with pUR2434, and pUR2435 with pUR2436). This effect might be related to -not detected- copy number variations, plasmid stability effects, or transcriptional interference with other transcription processes on the plasmid in orientation I.

trans-formant	OD 550	α-gal total [mg/L]	secreted [%]	copy no.
pUR2433-1	10.7	50.4	99,7	20
-2	10.0	35,0	99.6	20
pUR2434-1	9.5	114.2	99.7	15
-2	11,7	149,6	99.7	20
pUR2435-1	10.5	28,8	90.5	20
-2	10.1	24.7	91.5	20
pUR2436-1	11.1	82.7	83.1	20
-2	10.6	58.4	84.3	10
pUR2437-1	9.9	0.23	83.6	1
-2	11.4	0.21	90.5	1
-3	10.5	0.14	71.4	1
pUR2438-1	11.0	0.09	77.8	1
-2	11.0	0.12	75.0	1
-3	10.5	0.14	85.7	1

The expression of the integration vectors (pUR2437 and pUR2438) was very low, a result, which could be correlated to the low copy number present in the cell. One possible explanation for this effect is the presence of the intact heterologous, *S. cerevisiae LEU2* marker gene on the cassette; since this promoter might be strong enough to supply the leu deficient cell with sufficient gene product, even in this single copy configuration.

By taking benefit of the very recently cloned and sequenced *LEU2* gene of *K. marxianus* (Bergkamp *et al.*,1991), one can further enhance stability and expression of homologous and

heterologous genes in the described *K. marxianus* strain KMS1, by replacing the *LEU2* copy from *S. cerevisiae* in pUR2437 and pUR2438 with the corresponding homologous *LEU2* and *LEU2d* promoter deficient gene copies from *K. marxianus* on the integration cassette.
Moreover, the cloned *LEU2* gene can be further used to obtain a disruption/deletion *leu2* auxotrophic mutant strain without insertion of heterologous DNA.

Example 13. Construction of KMS3

For the construction of a strictly homologous *K. marxianus leu2* mutant the *URA3* gene of strain CBS 6556 was isolated first and subsequently utilized for the disruption of the *LEU2* locus.
The last step in the biosynthesis of pyrimidine is catalysed by orotidine-5' phosphate carboxylase (EC 4.1.1.23), an enzyme which in *S. cerevisiae* is encoded by the *URA3* gene. The *URA3* gene is one of the most commonly used selection markers because of the availability of counter selection for the marker (Boeke et al., 1984). Yeast cells having an active *URA3* gene are unable to grow in medium containing 5-fluoro-orotic acid (5-FOA), while ura3 mutants grow normally.
The *URA3* gene of *K. marxianus* CBS 6556 was isolated by screening a genomic *K. marxianus* DNA bank (Bergkamp et al., 1991), inserted in the vector lambdaL47.1 (Loenen et al., 1980) with a radioactively labelled *S. cerevisiae URA3* DNA fragment. Three phage clones, which hybridized with the *URA3* probe were isolated by standard techniques.
Restriction analysis followed by Southern analysis in all cases detected a 2.5 Kbp *Eco*RI/*Sph*I fragment which was subcloned in pUC19, resulting in plasmid pKMU1.
This insert carried by this plasmid and several subclones thereof were sequenced by described methods; the DNA sequence and the corresponding amino acid translation are given in Figure 12 (corresponding to Seq. ID. No. 11 and 12). The determined DNA sequences showed 71% homology on

the DNA level and 81% homology on the amino acid level with the corresponding *S. cerevisiae URA3* gene and the product (Rose et al., 1984).

For the construction of the food-grade *leu2* mutant spontaneous ura- mutants were selected by plating *K. marxianus* wt cells on 5-FOA plates. Out of 10^8 cells, 4 uracil requiring mutants were obtained; one of these -named KMS2- was used for further construction work.

Plasmid pKMU2 was constructed by replacing parts of the 5.1 kb *Eco*RI *LEU2* fragment in plasmid pKML1 (Bergkamp et al., 1991) with the *K. marxianus URA3* gene as indicated in Figure 13. A 1 kb *Stu*I/*Eco*RV fragment containing a large part of the coding sequence of the *LEU2* gene was replaced by a 2.5 kb *Eco*RI/*Sph*I fragment containing the *K. marxianus URA3* gene, giving plasmid pKMU2. *Eco*RI and *Sph*I sticky ends were made blunt by use of T4 DNA polymerase in the presence of all four dNTP's The linear 6.5 kb *Eco*RI fragment, containing the *leu2* gene disruption *leu2::URA3* was further used to transform KMS2 by electroporation as described by Meilhoc et al. 1990 with selection on medium lacking uracil to select for uracil prototrophy. From the 75 transformants obtained, 12 also displayed leucine requirement, indicating that in 63 transformants heterologous recombination had occurred, these transformants were not investigated any further. The Southern analysis of 3 of the 12 leu^- transformants confirm that in all cases the wild type copy of the *LEU2* gene has been replaced by *leu2::URA3* fragment. One of the newly obtained *leu2* transformants, designated KMS3, was stable even during long term growth under non selective conditions. This non-reverting *K. marxianus leu2* strain is suitable for overexpression of homologous or heterologous proteins in, for example, the food industry.

Example 14. Single copy integration of the α-galatosidase gene cassette into the INU1 locus

Multi-copy integration of a mRNA producing promoter-gene cassette into the rDNA locus generates an unusual DNA

arrangement consisting of sequences from the gene desired for expression which are transcribed by RNA polymerase II and the stable rRNA genes, transcribed by RNA polymerases I and III. To test the potential influence of the surrounding rDNA on the expression of α-galactosidase under the control of the INU promoter, this combination also tested at a different locus. The cassette was therefore integrated into the inulinase, *INU1*, locus through single cross over, thereby recombining the INU promoter with the wild type 5' upstream sequence.

For the construction of inulinase integration plasmids the *Cyamopsis tetragonoloba* α-galactosidase cassette with the described promoter fragment and the prepro-sequence of the *K. marxianus* inulinase (*INU1*) gene, and the *S. cerevisiae* PGK terminator, were combined in a plasmid incapable of replication in *K. marxianus*. To achieve this, the 804 bp long *Eco*RI/*Bsp*MI fragment of plasmid pUR2427 was ligated into the *Eco*RI/*Eag*I digested plasmid pSY9 (M. Harmsen, unpublished). The resulting plasmid, pUR2431 (Figure 9) was linearized with *Xho*I within the promotor sequence of the *INU1* gene prior to transformation to strain pUR2431 thereby preferentially targetting the integration into the chromosomal *INU1* locus. The expected integration event creates a chromosomal situation in which the α-galactosidase gene is placed under the regulation of the *INU1* promoter within the wt. chromosomal 5' DNA context, whereas the chromosomal *INU1* gene is placed under the control of the cloned *INU1* promoter fragment used in the fusion constructs (Figure 14.).

The acquired transformants were analyzed for both α-galactosidase- and inulinase production, α-galactosidase was measured as described earlier, while inulinase was measured as described by Rouwenhorst et al. (1988). Results for 4 different transformants are summarized below; the total α-galactosidase production and the extracellular inulinase activity of pUR2431 transformants and of the untransformed yeast strain KMS3 after growth for 24h in YPS

medium are given.

transformant/ strain	OD 660nm	α-galactosidase production (mg/l)	inulinase activity (U/ml)
IG1	7.8	0.02	150
IG2	8.0	0.02	189
IG3	7.2	0.2	233
IG4	8.7	18.3	13
KMS3	8.3	-	151

The transformant designated IG4 showed a strikingly high α-galactosidase production level compared to the other 3 transformants, concomitant with rather low inulinase production. Sothern analaysis of all 4 transformants revealed that only in transformant IG4 was plasmid pUR2431 integrated into the chromosomal inulinase locus in the expected manner, while in all three other cases the integration event occured elsewhere in the genome (not shown). Hence, the low inulinase production of this transformant might be caused by the lack of a further activating 5' DNA sequence element not present on the utilized fusion construct, an interpretation in accordance with the low expression results of the other transformants where the integration took place elsewhere and where the α-galactosidase gene is not under the control of the *INU1* promoter and further 5' upstream chromosomal inulinase sequences.

Example 15. Cloning of further *INU1* upstream sequences.
To obtain additional upstream sequences of the *INU1* promoter, which may enhance the expression directed by the

described *INU1* promoter, total chromosomal DNA of *K. marxianus* was isolated as described by Struhl et al. (1979) and digested with different restriction endonucleases, all having a recognition sequence within the first 500 bps of the cloned and sequenced DNA fragment containing the *INU1* locus. The fragments were separated by electrophoresis and the agarose gel subsequently subjected to Sothern analysis. For the identification of additional 5' sequences, plasmid pUR2421 was digested with *Eco*RI and *Nco*I and the about 290 bp long DNA fragment containing 5' sequence of the cloned promoter was isolated and used for the synthesis of a digoxygenin labelled DNA fragment, which was subsequently used as a DNA probe.

From the obtained specifically hybridizing signals, the *Eag*I digestion product with an apparent lenght of about 1.9 kb and the *Xho*I digestion product with a length of approximately 1 kb were chosen for further cloning. The *Eag*I digested *K. marxianus* DNA was separated by electrophoresis and fragments in the region of 1.9 kb purified from the gel by the procedure described earlier. These fragments were ligated into *Eag*I digested Bluescript vector (Stratagene) and the products introduced by transformation into *E. coli* JM109. The transformants were secreened by colony hybridization as described earlier using the DIG labelled DNA probe containing 5' sequences from the *INU1* promoter described above. A similar approach was used to clone the *Xho*I fragment containing upstream sequences from *INU1* but in this case the vector was digested with *Xho*I prior to ligation with *Xho*I digested *K. marxianus* DNA of approximately 1 Kb in size.

Using these techniques plasmids pUR2440 and pUR2439 were identified which contain the approximately 1.9 kb *Eag*I fragment and the approximately 1.1 Kb *Xho*I fragments of the *INU1* promoter respectively (Figure 15). The DNA sequence of an approximately 470 bp region immediately 5' to the previously cloned *INU1* sequences was determined by the

techniques described earlier and is shown in figure 16 (corresponding to Seq. ID. No. 13). Plasmid pUR2440 was deposited in the Centraalbureau voor Schimmelcultures, Baarn, The Netherlands under accession number CBS 648.93.

Example 16. Construction of expression vectors containing longer derivatives of the *INU1* promoter.

To obtain an autonomously replicating vector carrying the extended *INU1* promoter sequence plasmid pUR2434 was partially digested with *Eco*RI and the approximately 11.6 Kb linear fragment isolated from a gel and dephosphorylated. The approximately 470 bp *Eco*RI fragment from pUR2440 was subsequently ligated into this vector. The ligation mix was used to transform *E. coli* JM109 to ampicillin resistance and the plasmid DNA from the resulting transformants analysed by digestion with *Nco*I to identify those which contained the insert adjacent to the existing *INU1* promoter sequences in pUR2434. Sequencing across the relevant *Eco*RI site in pUR2440 using primer INUT (Figure 18) which hybridizes within the previously sequenced region of the *INU1* promoter was carried out to determine the sequence of the extended promoter. The orientations of the cloned approximately 470 bp *Eco*RI fragments were confirmed by sequencing using primer INUT. Two plasmids were identified which carried the fragment in the correct orientation and 4 were found with the fragment in the opposite orientation. A still longer upstream region can be cloned into pUR2434 by cloning of the 1.4 kb fragment from pUR2440 produced by partial digestion with *Eco*RI into pUR2434 partially digested with *Eco*RI as described above. The correct orientation of the fragment can be easily determined by digestion of the resulting plasmids with *Eag*I which will release the approximately 1.9 kb fragment cloned in pUR2440.

Plasmid pUR2445 (Figure 18) carrying the extended promoter and pUR2434 were introduced by electroporation (Bolen *et al.*, 1990) into *K. marxianus* strain KMS3 with selection for

leucine prototrophy. Representatives of the resulting transformants were grown overnight in minimal medium at 37 °C and diluted 1:10 into 10 ml of YEP, 5% sucrose induction medium. These cultures were grown for a further 24 hours at 37°C and the α-galactosidase levels in the fermentation meduim determined. The results are shown below:

TRANSFORMANT STRAIN	OD_{660}	α-GALACTOSIDASE (UNITS)
KMS3 pUR2434 (1)	24	7
KMS3 pUR2434 (2)	22	10
KMS3 pUR2434 (3)	24	11
KMS3 pUR2445 (1)	27	24
KMS3 pUR2445 (2)	19	18
KMS3 pUR2445 (3)	27	21

From these results it appears as if the extension of the promoter has a beneficial effect upon enzyme production levels.

Multi-copy rDNA integrative plasmids carrying a longer *INU1* promoter can similarly be constructed using the 1.9 kb *Eag*I fragment from pUR2440. The integrative vector pUR2437 can be partially digested with *Eco*RI and either the approximately 470 bp *Eco*RI fragment from pUR2440 or 1.4 kb fragment formed by partial digestion of pUR2440 with *Eco*RI inserted upstream of the existing *INU1* sequences. The resulting ligation mix can be introduced into *E.coli* JM109 and plasmids containing inserts identified by digestion with *Eco*RI. The orientation of the approximately 470 bp fragment could be confirmed by sequencing and that of the approximately 1.4 kb *Eco*RI fragment by digestion of the plasmids with *Eag*I which should give a fragment of approximately 1.9 kb in addition to those derived from the vector. The plasmids so produced can be linearized by

digestion with *Xba*I and introduced by transformantion into *K. marxianus* strain KMS1 or KMS3 with selection for leucine prototrophy.

Literature References

Berg, v/d J.H., Laken, v/d K.J., Ooyen, A.J.J., Renniers, T., Rietveld, K., Schaap, A., Brake, A.J., Bishop, R.J., Schultz, K., Moyer, D., Richmann, M, Shuster, J., 1990. *Kluyveromyces* as a host for heterologous gene expression: expression and secretion of prochymosin. Biotech. **8**, 135-139

Bergkamp, R.J.M., Geerse, R.H., Verbakel, J.M.A., Musters, W., Planta, R.J., 1991. Cloning and disruption of the *LEU2* gene of *Kluyveromyces marxianus* CBS 6556. Yeast, Vol **7**, 963-970.

Bergkamp, R.J.M., Kool, I.M., Geerse, R.H., Planta, R.J., 1992. Multiple copy integration of the α-galactosidase gene from *Cyamopsis tetragonoloba* into the ribosomal DNA of *Kluyveromyces lactis*. Curr. Genet **21**, 365-370.

Bianchi, M.M., Falcone, C., Chen, X.J., Wésolowski-Louvel, M., Frontali, L., and Fukuhara, H., 1987. Transformation of the yeast *Kluyveromyces lactis* with new vectors derived from the 1.6 μm circular plasmid pKD1. Cur. Genetics **12**: 185-192.

Boeke J.D., LaCroute, F., Fink, G.R. (1984). A positive selection system for mutants lacking orotidine -5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen. Genet. 197, 345-346.

Bolen, P.L., and McCutchan, J.E., 1990. Electroporation of *Kluyveromyces lactis*. **In** Mager, W.H., and Planta, R.J. (eds.).15th International conference on yeast genetics and

molecular biology, The Hague, Netherlands. Wiley, Chichester.

Carter, B.L.A., Irani, M., Mackay, V.I., Seale, R.L., Sledziewski, A.V., Smith, R.A., 1988. Expression and secretion of foreign genes in yeast. In: Glover, D.M., (Ed.) DNA Cloning III. A practical approach. IRL press Oxford. 141-161.

Chen, X.J., Bianchi, M.M., Suda, K., and Fukahara, H., 1989. The host range of pKD1 derived plasmids in yeast. Curr. Genet. **16**, 95-98.

Chung, C.T., Niemela, S,L., Miller, R.H., 1989. One-step preparation of competent *E.coli*: Transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. USA, **86**; 2172-2175

Crowe, J.S., Cooper, H.J., Smith, M.A., Sims, M.J., Parker, D., and Gewert, D., 1991. Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinaseK digestion. Nucl. Acids Res. Vol. 19, p. 184.

Falcone, C., Saliola, M., Chen, X.J., Frontali, L., Fukuhara, H., 1986. Analysis of a 1.6-μm circular plasmid from the yeast *Kluyveromyces drosophilarum*: structure and molecular dimorphism. Plasmid 15: 248-252

Fleer, R., Yeh, P., Amellal, N., Maury, I., Fournier, A., Bacchetta, F., Baduel, G., Jung, G., Hote, H.L., Becquart J., Fukuhara H., Mayaux, J.F., 1991. Stable multicopy vectors for high-level secretion of recombinant human serum albumin by *Kluyveromyces* yeasts. Biotech., Vol 9, 969-975

Fuller, R.S., Sterne, R.E., Thorner J., 1988. Enzymes required for yeast prohormone processing. Ann. Rev. Physiol. **50**, 345-362

Giuseppin, M.L.F., Lopes, M.T.S., Planta, R.J., Verbakel, J.M.A., Verrips, C.T., 1991. Process for preparing a protein by a fungus transformed by multicopy integration of an expression vector. PCT International WO 91/00920

Hamilton, R., Watanabe, K.C., de Boer H.A., 1987. Compilation and comparison of the sequence context around the AUG start codons in *S. cerevisiae* mRNA's. Nucl. Acids. Res. **15,** 3381-3593.

Heijne, van, G., 1986. A new method for preceding signal sequence cleavage sites. Nucl. Acids Res. **16**: 4683-4690.

Hsiao, K., 1991. A fast and simple procedure for sequencing double stranded DNA with Sequenase. Nucl. Acids Res. **19:** 2787.

Köhrer, K., and Domdey, H., 1991. Preparation of high molecular weight RNA, pp. 398-405. **In** Guthrie, C., Fink, G.R. (eds.), Methods in Enzymology **194:** Guide to yeast genetics and molecular biology, Academic Press, Inc., Harcourt Brace Jovanovich, San Diego.

Laloux, O., Cassert, J.P., Delcour, J., Van Beeumen, J., Vandenhaute, J., 1991. Cloning and sequencing of the inulinase gene of *Kluyveromyces marxianus* var. *marxianus* ATCC 12424. FEBS lett. ,**289**, 64-68.

Loenen, W.A.M., Brammar, W.J., 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction sites. Gene **20**, 249-259.

Lopes, T.A., Klootwijk, J., Venstra, A.E., van der Aar, P.C., van Heerikhuizen, H., Raue, H.A., Planta, R.J., 1989. High-copy-number integration into the ribosomal DNA of *S. cerevisiae*: a new vector for high level expression. Gene **79,** 199-206

Mead, D.A., Szczesna-Skorupa, E., and Kemper, B., 1986. Single-stranded 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein engineering **1**: 67-74.

Meilhoc, E., Masson, J., Teissiè, J., (1990). High efficiency transformation of intact yeast cells by electric field pulses. Bio/Technology **8**, 223-227.

Nehlin, J.O., Ronne, H., 1990. Yeast MIG1 repressor is related to the mammalian early growth response and Wilms' tumor finger proteins. EMBO, **9**, 2891-2898.

Nehlin, J.O., Carlberg, M., Ronne, H., 1991. Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. EMBO, **10**, 3373-3377.

Rosè, M., Grisafi, P., Botstein, D. (1984). Structure and function of the yeast *URA3* gene: expression in *E. coli*. Gene, **29**, 113-124.

Rouwenhorst, R.J., Visser, L.E., van der Baan, Scheffers, W.A., van Dijken, J.P. (1988) Production, distribution and kinetic properties of inulinase in continous culture of *Kluyveromyces marxianus* CBS 6556. Appl. Environm. Microbiol. 54; 1131-1137.

Rouwenhorst, R.J., Hensing, M., Verbakel, J., Scheffers, W.A., and van Dijken, J.P., 1990. Structure and properties of the extracellular inulinase of *Kluyveromyces marxianus* CBS 6556. Appl. Environ. Microbiol. **56**: 3337-3345.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning. A laboratory manual. Second edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sanger, F., Nicklen, S., and Coulson, A.R., 1977. DNA

sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463-5467.

Struhl, K., Stinchcomb, D.T., Scherer, S., and Davis., R.W., 1979. High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. Proc. Natl. Acad. Sci. USA. **76**: 1035-1039.

Szybalski,W., Kim, S.C., Hasan, N., Podhajska, A.J., 1991. Class-IIS restriction enzymes: a review. Gene **100**, 13-26.

Tanguy-Rougeau, C., Weselowsky-Louvel, M., Fukuhara., H., 1988. The *Kluyveromyces* KEX1 gene encodes a subtilisin-type serine protease. FEBS Lett. **234**, 464-470.

Vandamme, E.J., Derycke, D.G., 1983. Microbiol. inulinases. Fermentation process; properties and applications. Adv. Appl. Microbiol. **29**, 139-176

Verbakel, J.M.A., 1991. Heterologous gene expression in the yeast Saccharomyces cerevisiae. Phd thesis.

Yanisch-Perron, C., Vieira, J., and Messing, J., 1975. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC vectors. Gene **33**: 103-119.

Samples of *Escherichia coli* transformed with plasmids pUR2421 and pUR2422 were deposited under the Budapest Treaty at the Centraalbureau for Schimmelcultures (CBS) in Baarn, The Netherlands on 27 May 1992. They received deposit numbers CBS 265.92 and CBS 266.92, respectively. These plasmids were mentioned in the specification in Examples 5 and 6 and in Example 8. Plasmid pUR2440 (example 15) was also deposited at the CBS under accession number CBS 648.93.

C L A I M S

1. A nucleic acid sequence derivable from a yeast and comprising at least a regulatory region derivable from a gene encoding a polypeptide having inulinase activity or a modified sequence of said nucleic acid sequence also having regulatory activity.

2. A nucleic acid sequence according to claim 1, wherein said yeast belongs to the genus *Kluyveromyces*, preferably belonging to the species *Kluyveromyces marxianus*, more preferably said yeast being the strain *K. marxianus* var. *marxianus* deposited at the CBS at Baarn, The Netherlands under number CBS 6556.

3. A nucleic acid sequence according to any of the previous claims, comprising at least one region selected from the group consisting of a promoter, a termination signal, and a sequence encoding a secretory signal necessary for secreting a gene product from a yeast, the latter preferably being derivable from the inulinase gene of a *Kluyveromyces* yeast.

4. A nucleic acid sequence derivable from a yeast and comprising at least one regulatory region derivable from a gene encoding a polypeptide having inulinase activity according to any of the previous claims, comprising at least a part of the nucleic acid sequence of Figure 5 or an equivalent nucleic acid sequence, preferably comprising at least polynucleotide -737 to -1 of Figure 5 having promoter activity or comprising at least polynucleotide 1 to 48 of Figure 5 encoding the inulinase pre-sequence.

5. A recombinant nucleic acid sequence according to any of the previous claims, wherein the regulatory region is operably linked to DNA encoding a specific RNA sequence not coding for a specific protein.

6. A vector comprising a nucleic acid sequence according to any of the claims 1-4, said nucleic acid sequence being operably linked to a structural gene, such as a gene encoding a polypeptide having inulinase activity or α-galactosidase activity, preferably a yeast vector.

7. A recombinant host cell comprising a nucleic acid sequence according to any of the claims 1-5 or a vector according to claim 6.

8. A process for producing a desired expression product, wherein a host cell comprising a vector according to claim 6 is cultured under conditions enabling the structural gene to be expressed and optionally the resulting desired expression product is isolated.

9. A process for producing RNA, wherein a host cell comprising a recombinant nucleic acid sequence according to claim 5 is cultured under conditions enabling production of the RNA, whereby preferably the resulting RNA influences the formation of at least one metabolite or the resulting RNA is a flavouring component.

10. Use of a part of a nucleic acid sequence according to any of claims 1-4 as a probe or a primer said part having a length of at least 10 nucleotides.

Fig.1. Internal sequences of inulinase after CNBr-digestion

fragment 1 from Inulinase forms I and II

```
                *               *
M-↓-V-I-D-Y-n-N-T-S-G-F-F-n-S-S-V-D-P-R-Q--r-A-V-A-V
```

fragment 2 from Inulinase forms I and II

↓-P-K-V-F-W-↓

Fig.2a. **DNA oligonucleotides derived from amino acid sequence analysis**

a. DNA probes from the CNBr fragments 1 and 2.

fragment 1:

	M-	**↓-V-**	**I-**	**D-**	**Y-**	**n-**	**N-**	**T-**	
	ATG	GTT	ATT	GAT	TAC	AAC	AAC	ACT	
		C	C	C	T	T	T	C	
		G	A					G	
		A						A	
3'	TAC	CAK	TAK	CTG	ATG	TTG	TTG	TG	5' [23] KLM 06 K=G/T

fragment 2:

	D-	**P-**	**K-**	**V-**	**F-**	**W-**	
	GAT	CCA	AAG	GTT	TTC	TGG	
	C	T	A	C	T		
		G		G			
		C		A			
3'	CTG	GGK	AAT	CAK	AAG	ACC	5' [18] KLM 07 K=G/T

Fig.2b. DNA probes from the N-terminal fragment

```
                K-  A-  I-  T-  G-  T-  T-  F-
               AAG GCT ATT ACT GGT ACT ACT TTC
                 A   A   C   C   A   C   C   T
                     C   A   A   C   A   A
                     G       G   G   G   G
          3'   TTT CGK TAK TGK CCK TGK TGK AAG  5' [24] KLM
04 K=G/T
          3'   TTT CGR TAR TGR CCR TGR TG       5' [20] KLM
05 R=G/A
          5'   AAG GCK ATK ACK GGK ACK ACK TTT  3' [24] KLM
08 K=G/T
          5'   AAG GCY ATY ACY GGY ACY AC       3' [20] KLM
09 Y=T/C
```

Figure 3: Inulinase PCR fragment in pUR 2415.

seq. = a.a. sequence determined with amino acid sequencing

Underlined amino acids are identical to Saccharomyces cerevisiae invertase

Figure 4: Cleavage map.

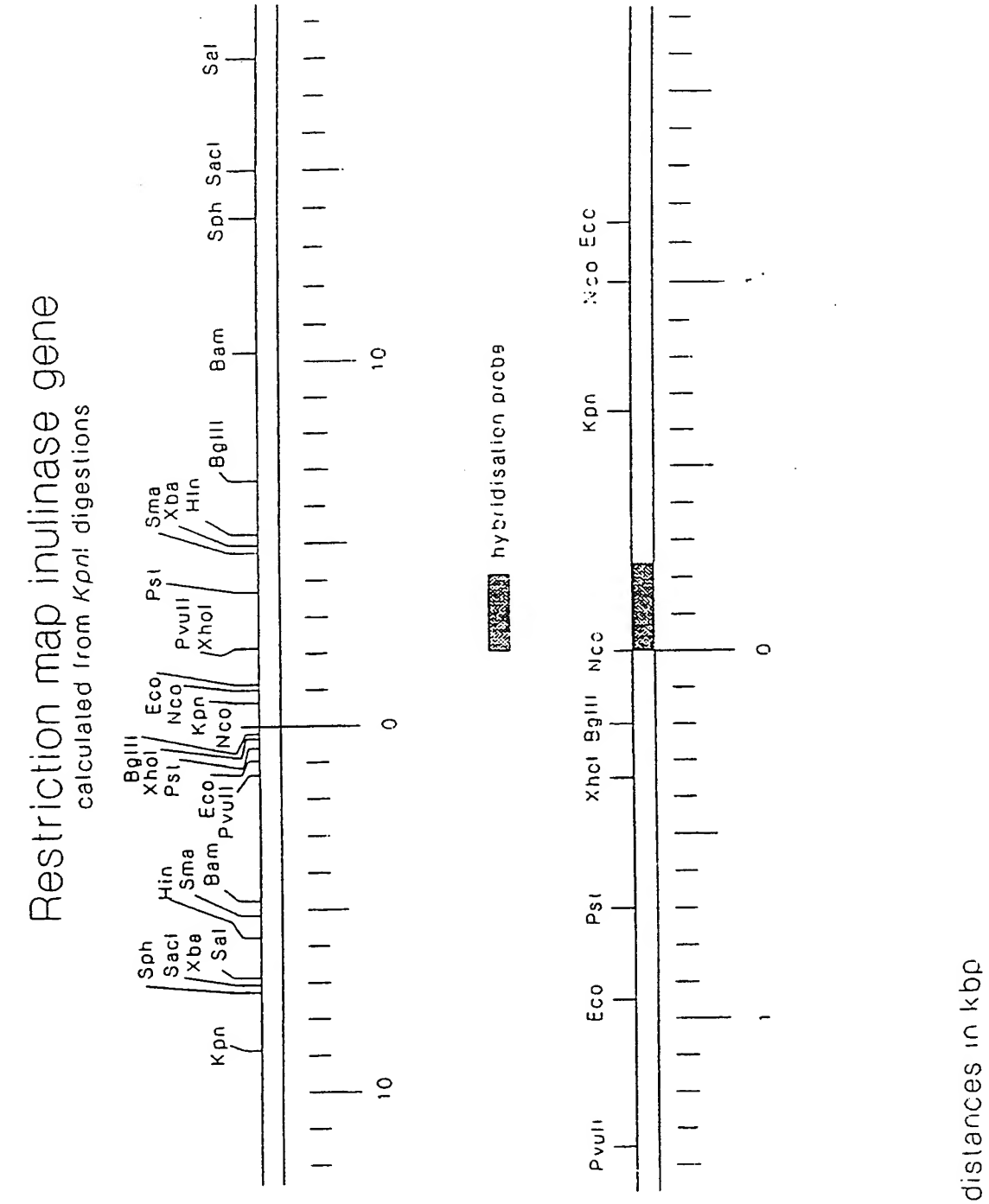


Figure 5: Nucleotide sequence of the inulinase gene (corresponding to Seq. ID. No. 9.

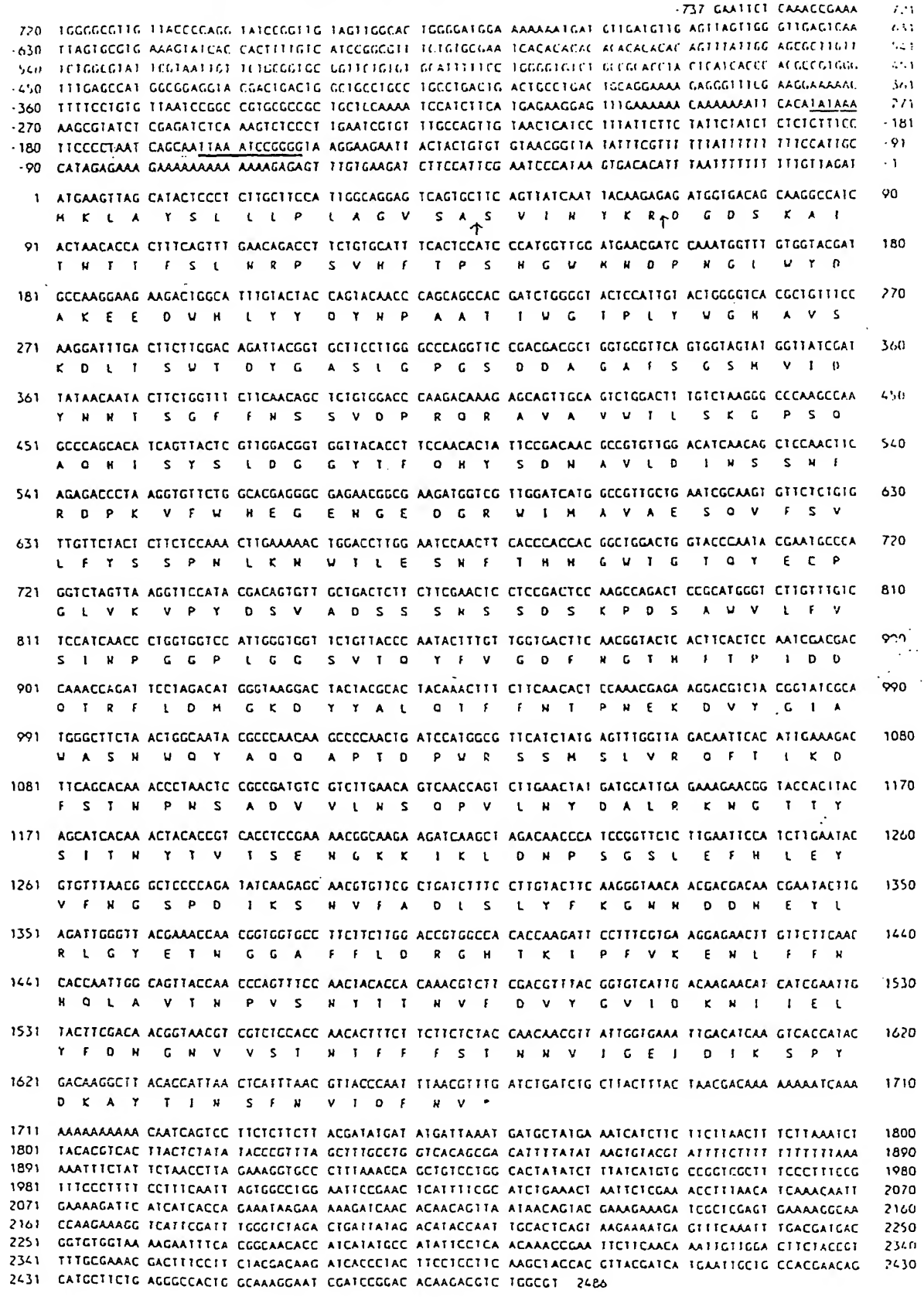
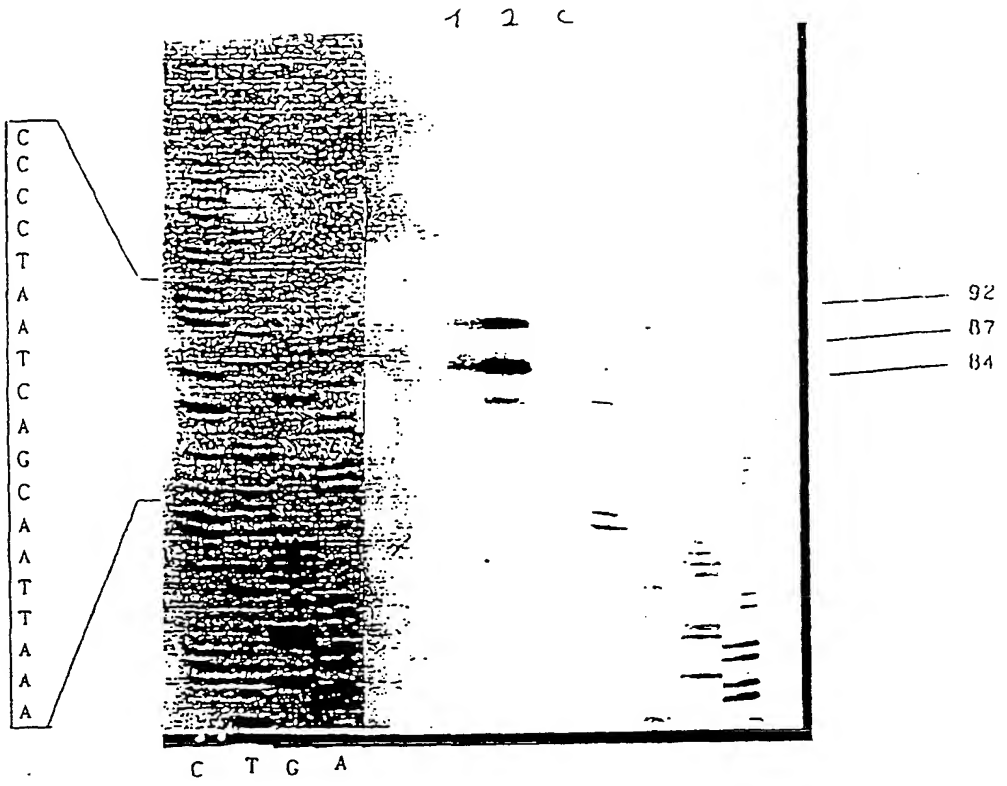


Figure 6.

A: Hybridisation with primer p16



B: Hybridisation with primer p21

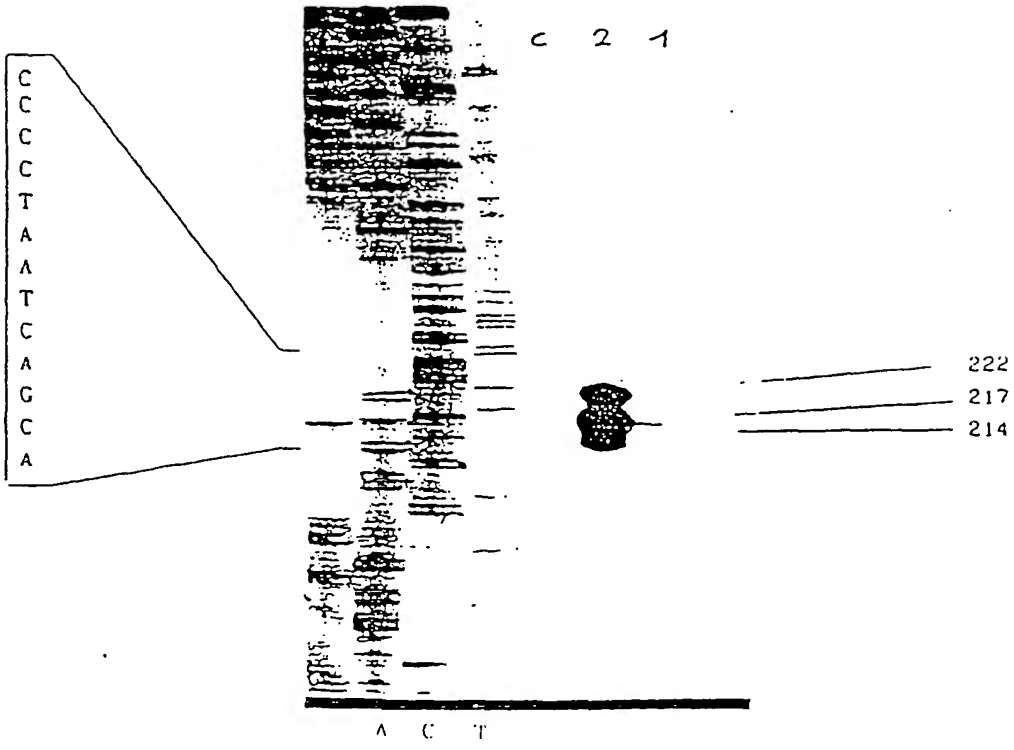


Figure 7.

a) Invertase signal sequence - α-galactosidase linkage:

```
<inv. ss< > α-gal >                          > α-gal >
    A   A   E   N                 EagI         A   E   N
  T GCG GCC GAA AAC               --->     G GCC GAA AAC
  A CGC CGG CTT TTG                              CTT TTG
```

b) Inulinase prepro-link

```
          ↓ sign pept.              ↓ mature inulinase
    S  A  S  V  I  N  Y  K  R  D  G  D  S  K  A  I  T  N
<-TCAGTGCTTCAGTTATCAATTACAAGAGAGATGGTGACAGCAAGGCCATCACTAAC->
        3'-GTCAATAGTTAATGTTCTCCCCGGGTACCGTCCATTCGAACCC-5'
```

↓ PCR

```
    S  A  S  V  I  N  Y  K  R ApaI NcoI      HindIII
<-TCAGTGCTTCAGTTATCAATTACAAGAGGGCCCATGGCAGGTAAGCTTGGG
<-AGTCACGAAGTCAATAGTTAATGTTCTCCCGGGTACCGTCCATTCGAACCC
                                    BspMI
```

↓ BspMI

```
    S  A  S  V  I  N  Y  K  R
<-TCAGTGCTTCAGTTATCAATTACAAGAG
<-AGTCACGAAGTCAATAGTTAATGTTCTCCCGG
```

c) Inulinase pre-link

```
                                  ↓ signal pept.        ↓ mat.
   L  L  L  P  L  A  G  V  S  A  S  V  I  N  Y  K  R  D  G
<-CCTCTTGCTTCCATTGGCAGGAGTCAGTGCTTCAGTTATCAATTACAAGAGAGATGGT-
>          3'-GTAACCGTCCTCAGTCACGCCCGGGCGTACGTCCATTCGAACCC-5'
```

↓ PCR

```
   L  L  L  P  L  A  G  V  S  A NotI SphI     HindIII
<-CCTCTTGCTTCCATTGGCAGGAGTCAGTGCGGCCGCATGCAGGTAAGCTTGGG
<-GGAGAACGAAGGTAACCGTCCTCAGTCACGCCGGCGTACGTCCATTCGAACCC
                                       BspMI
```

↓ BspMI (or NotI)

```
   L  L  L  P  L  A  G  V  S  A
<-CCTCTTGCTTCCATTGGCAGGAGTCAGTGC
<-GGAGAACGAAGGTAACCGTCCTCAGTCACGCCGG
```

Figure 8.

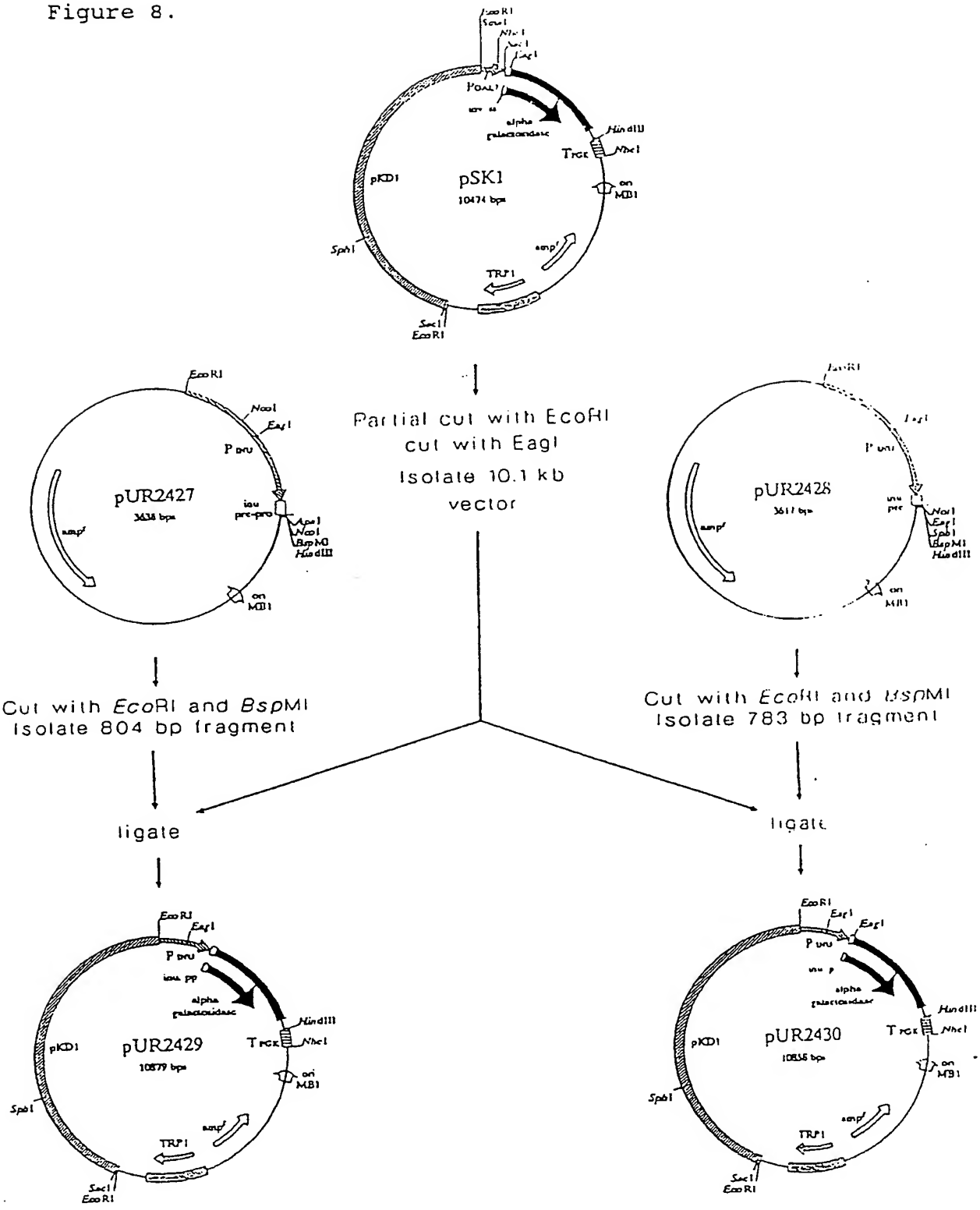

pSK1
pKD1
TRP1
amp'
Partial cut with EcoRI
cut with EagI
Isolate 10.1 kb
vector
pUR2427
pUR2428
Cut with EcoRI and BspMI
Isolate 804 bp fragment
Cut with EcoRI and BspMI
Isolate 783 bp fragment
ligate
ligate
pUR2429
pUR2430

Figure 9.

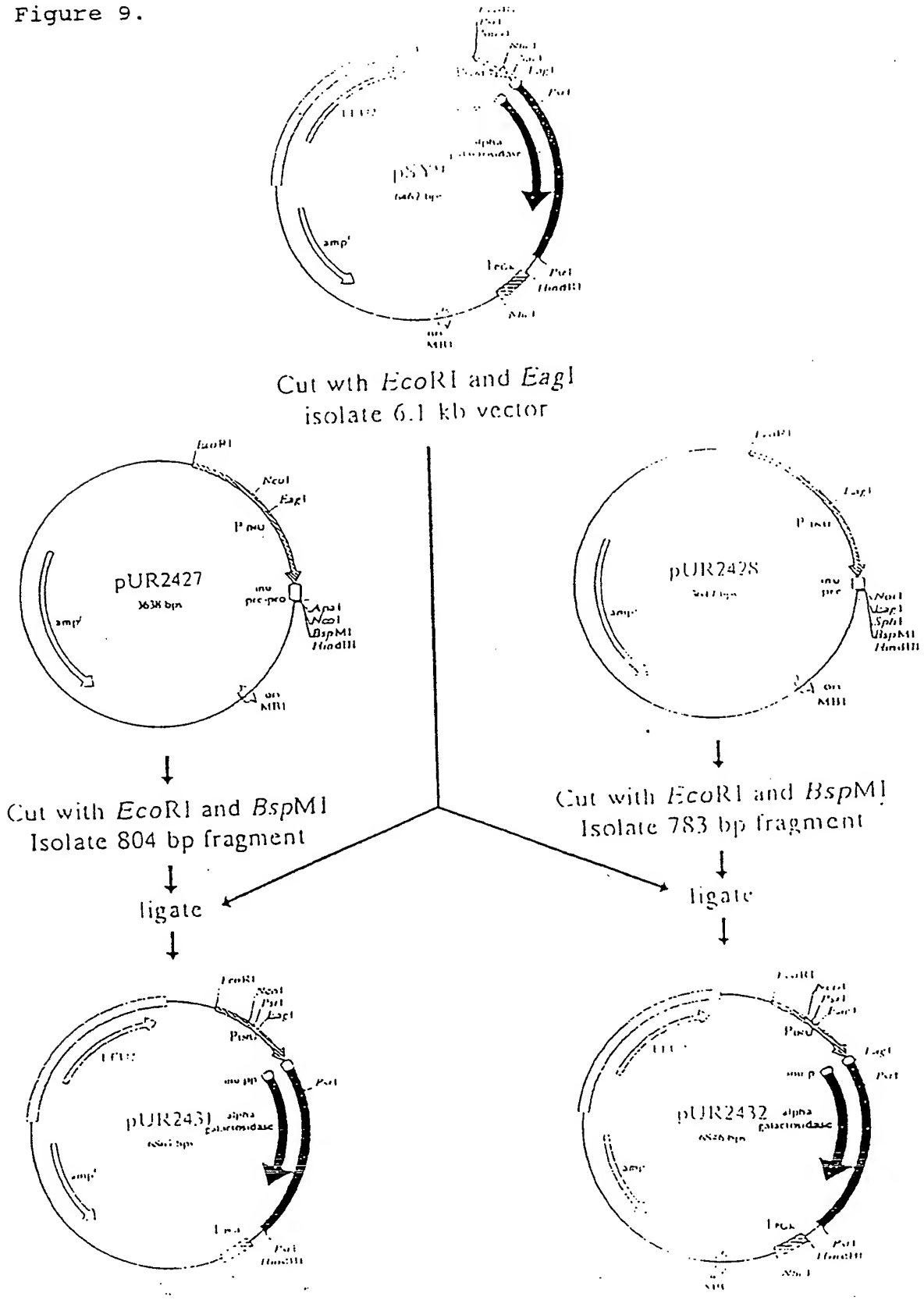

Cut wth EcoRI and EagI
isolate 6.1 kb vector
pUR2427
3638 bp
Cut with EcoRI and BspMI
Isolate 804 bp fragment
ligate
pUR2428
Cut with EcoRI and BspMI
Isolate 783 bp fragment
ligate
pUR2431
pUR2432

Figure 10.

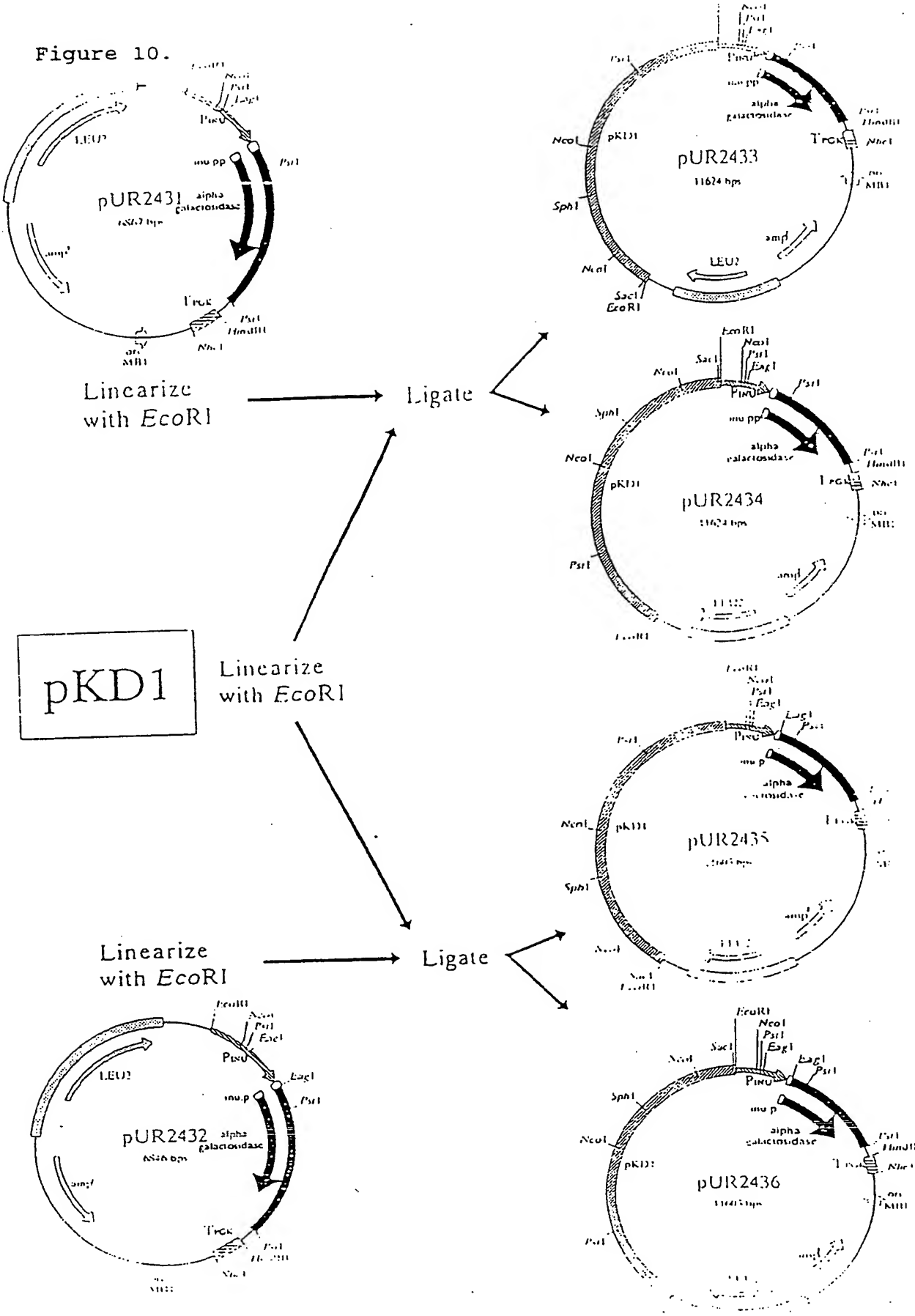

pUR2431
pUR2433
pUR2434
pUR2435
pUR2436
pUR2432
pKD1
Linearize with EcoRI
Ligate
Linearize with EcoRI
Ligate
LEU2
alpha galactosidase
TPGK
pKD1
EcoRI
NcoI
PstI
EagI
SphI
SacI
HindIII
NheI

Figure 11.

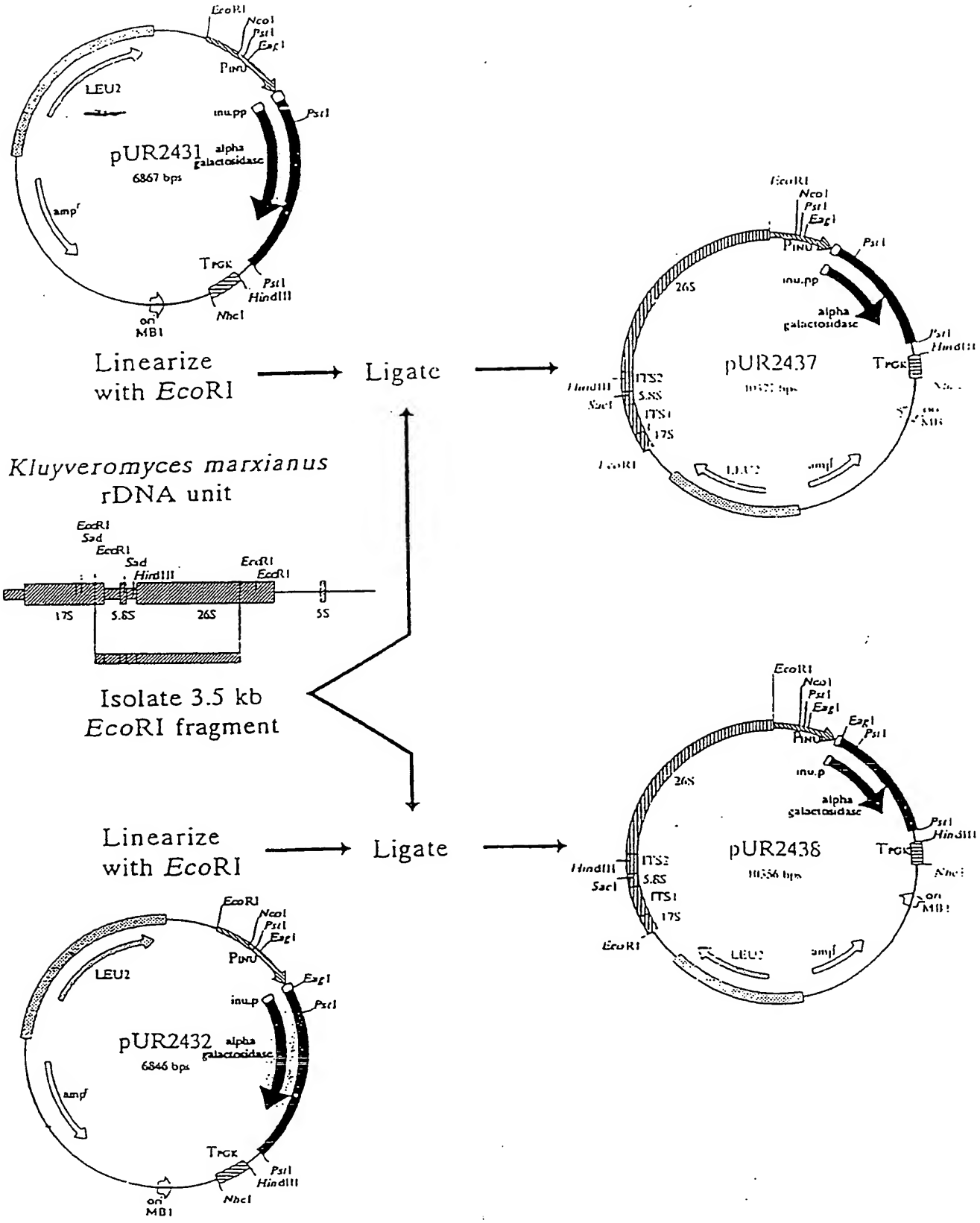

pUR2431
6867 bps
LEU2
inu.pp
alpha galactosidase
amp
TPGK
PstI
HindIII
NheI
ori MB1
Linearize with EcoRI
Ligate
pUR2437
Kluyveromyces marxianus rDNA unit
EcoRI
SacI
SacI
HindIII
EcoRI
EcoRI
17S
5.8S
26S
5S
Isolate 3.5 kb EcoRI fragment
Linearize with EcoRI
Ligate
pUR2438
pUR2432
6846 bps

Figure 12.

```
                GAGCA TCTTGGTCTT CTGAGCTCAT TATACCTCAA TCAAAACTGA AATTAGGTGC    -491
CTGTCACGGC TCTTTTTTTA CTGTACCTGT GACTTCCTTT CTTATTTCCA AGGATGCTCA TCACAATACG    -421
CTTCTAGATC TATTATGCAT TATAATTAAT AGTTGTAGCT ACAAAAGGTA AAAGAAAGTC CGGGGCAGGC    -351
AACAATAGAA ATCGGCAAAA AAAACTACAG AAATACTAAG AGCTTCTTCC CATTCAGTCA TCGCATTTCG    -281
AAACAAGAGG GGAATGGCTC TGGCTAGGGA ACTAACCACC ATCGCCTGAC TCTATGCACT AACCACGTGA    -211
CTACATATAT GTGATCGTTT TAACATTTTC AAAGGCTGTG TGTCTGGCTG TTTCCATTAA TTTTCACTGA    -141
TTAAGCAGTC ATATTGAATC TGAGCTCATC ACCAACAAGA AATACTACCG TAAAAGTGTA AAAGTTCGTT     -71
TAAATCATTT GTAAACTGGA ACAGCAAGAG GAAGTATCAT CAGCTAGCCC CATAAACTAA TCAAAGGAGG      -1

ATG TCG ACT AAG AGT TAC TCG GAA AGA GCA GCT GCT CAT AGA AGT CCA GTT GCT GCC AAG     60
Met Ser Thr Lys Ser Tyr Ser Glu Arg Ala Ala Ala His Arg Ser Pro Val Ala Ala Lys    20

CTT TTA AAC TTG ATG GAA GAG AAG AAG TCA AAC TTA TGT GCT TCT CTT GAT GTT CGT AAA    120
Leu Leu Asn Leu Met Glu Glu Lys Lys Ser Asn Leu Cys Ala Ser Leu Asp Val Arg Lys    40

ACA GCA GAG TTG TTA AGA TTA GTT GAG GTT TTG GGT CCA TAT ATC TGT CTA TTG AAG ACA    180
Thr Ala Glu Leu Leu Arg Leu Val Glu Val Leu Gly Pro Tyr Ile Cys Leu Leu Lys Thr    60

CAT GTA GAT ATC TTG GAG GAT TTC AGC TTT GAG AAT ACC ATT GTG CCG TTG AAG CAA TTA    240
His Val Asp Ile Leu Glu Asp Phe Ser Phe Glu Asn Thr Ile Val Pro Leu Lys Gln Leu    80

GCA GAG AAA CAC AAG TTT TTG ATA TTT GAA GAC AGG AAG TTT GCC GAC ATT GGG AAC ACT    300
Ala Glu Lys His Lys Phe Leu Ile Phe Glu Asp Arg Lys Phe Ala Asp Ile Gly Asn Thr    100

  T AAA TTA CAA TAC ACG TCT GGT GTA TAC CGT ATC GCC GAA TGG TCT GAT ATC ACC AAT    360
Val Lys Leu Gln Tyr Thr Ser Gly Val Tyr Arg Ile Ala Glu Trp Ser Asp Ile Thr Asn    120

GCA CAC GGT GTG ACT GGT GCG GGC ATT GTT GCT GGT TTG AAG CAA GGT GCC GAG GAA GTT    420
Ala His Gly Val Thr Gly Ala Gly Ile Val Ala Gly Leu Lys Gln Gly Ala Glu Glu Val    140

ACG AAA GAA CCT AGA GGG TTG TTA ATG CTT GCC GAG TTA TCG TCC AAG GGG TCT CTA GCG    480
Thr Lys Glu Pro Arg Gly Leu Leu Met Leu Ala Glu Leu Ser Ser Lys Gly Ser Leu Ala    160

CAC GGT GAA TAC ACT CGT GGG ACC GTG GAA ATT GCC AAG AGT GAT AAG GAC TTT GTT ATT    540
His Gly Glu Tyr Thr Arg Gly Thr Val Glu Ile Ala Lys Ser Asp Lys Asp Phe Val Ile    180

GGA TTT ATT GCT CAA AAC GAT ATG GGT GGA AGA GAA GAG GGC TAC GAT TGG TTG ATC ATG    600
Gly Phe Ile Ala Gln Asn Asp Met Gly Gly Arg Glu Glu Gly Tyr Asp Trp Leu Ile Met    200

ACG CCA GGT GTT GGT CTT GAT GAC AAA GGT GAT GCT TTG GGA CAA CAA TAC AGA ACT GTG    660
Thr Pro Gly Val Gly Leu Asp Asp Lys Gly Asp Ala Leu Gly Gln Gln Tyr Arg Thr Val    220

GAT GAA GTT GTT GCC GGT GGA TCA GAC ATC ATT ATT GTT GGT AGA GGT CTT TTC GCA AAG    720
Asp Glu Val Val Ala Gly Gly Ser Asp Ile Ile Ile Val Gly Arg Gly Leu Phe Ala Lys    240

GGA AGA GAT CCT GTA GTG GAA GGT GAG AGA TAC AGA AAG GCG GGA TGG GAC GCT TAC TTG    780
Gly Arg Asp Pro Val Val Glu Gly Glu Arg Tyr Arg Lys Ala Gly Trp Asp Ala Tyr Leu    260

AAG AGA GTA GGC AGA TCC GCT TAA  GAGTTCTCCG AGAACATGCA GAGGTTCGAG TGTACTCGGA     844
Lys Arg Val Gly Arg Ser Ala End                                                  267

TCAGAAGTTA CAAGTTGATC GTTTATATAT AAACTATACA GAGATGTTAG AGTGTAATGG CATTGCGTCA     914
CATTGTATAC                                                                       924
```

Figure 13.

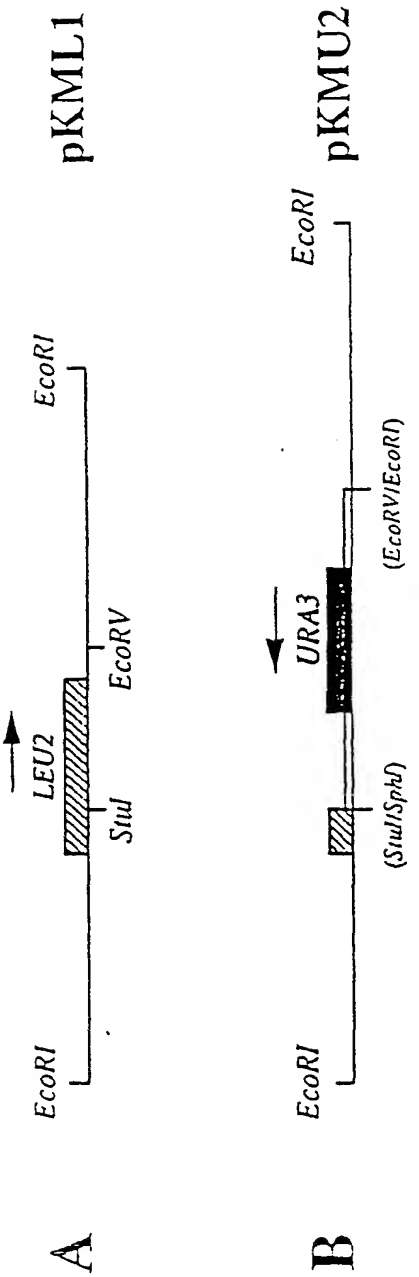

pKML1
EcoRI
LEU2
EcoRV
StuI
EcoRI
A
pKMU2
EcoRI
URA3
(EcoRV/EcoRI)
(StuI/SphI)
EcoRI
B

Figure 14.

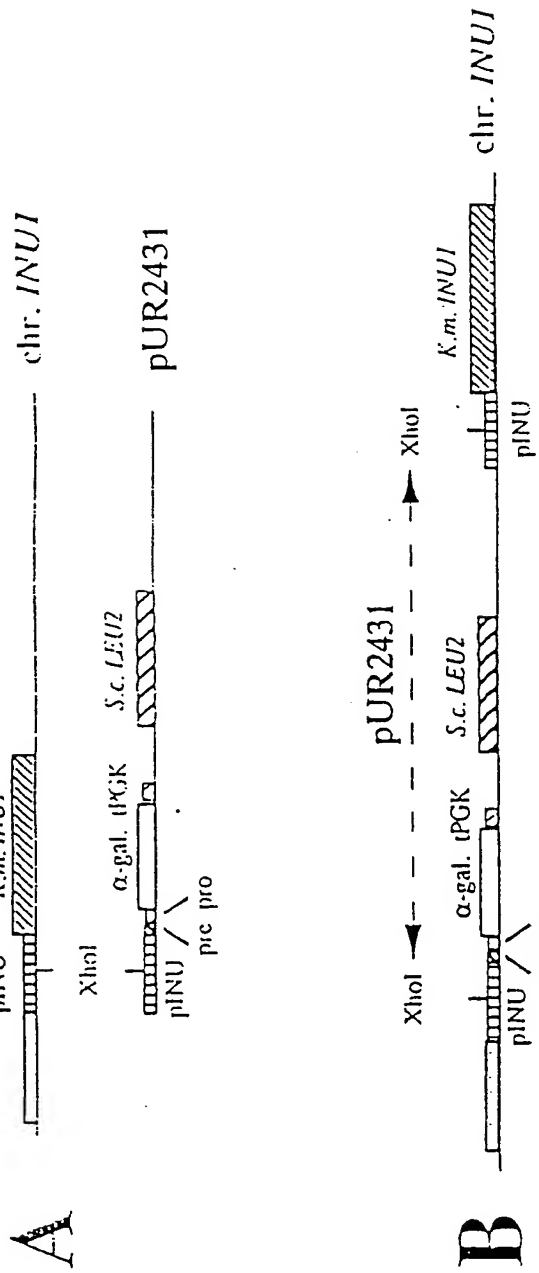

A
chr. INU1
K.m. INU1
pINU
XhoI
pINU
α-gal. tPGK
S.c. LEU2
pINU
pre pro
pUR2431
B
chr. INU1
K.m. INU1
pINU
XhoI
pUR2431
XhoI
α-gal. tPGK
S.c. LEU2
pINU
pre pro

Figure 15.

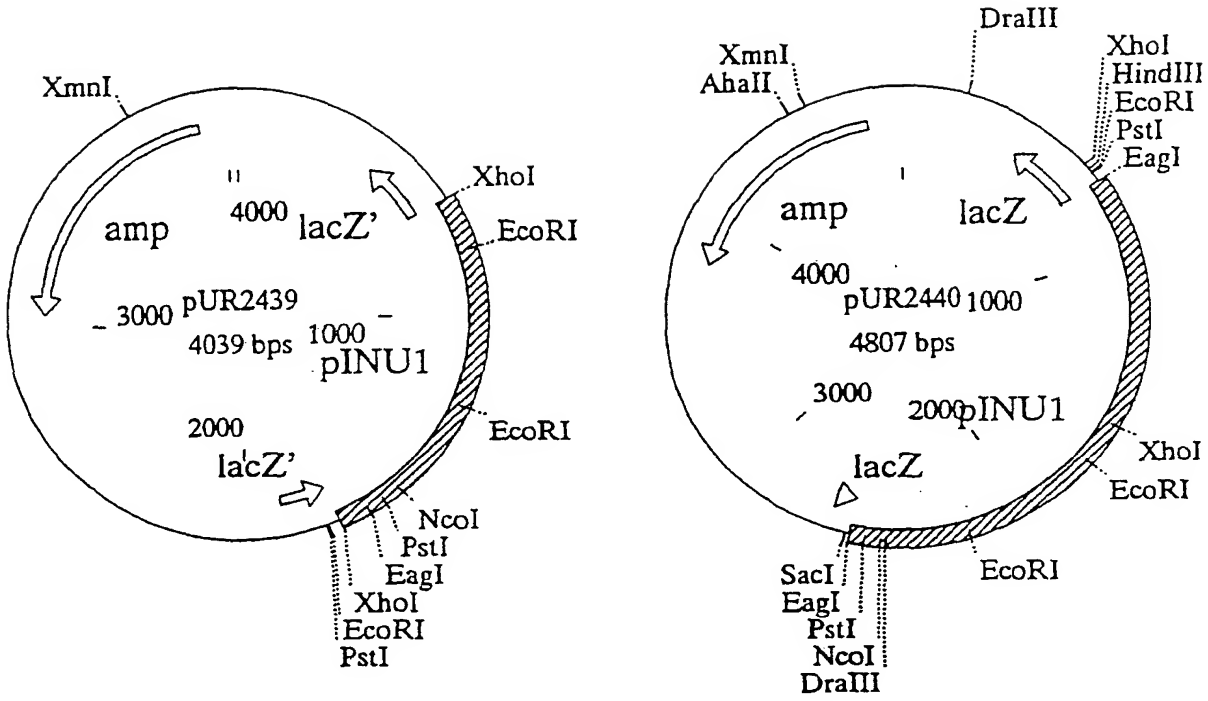

XmnI
amp
4000
lacZ'
XhoI
EcoRI
3000
pUR2439
4039 bps
1000
pINU1
2000
lacZ'
EcoRI
NcoI
PstI
EagI
XhoI
EcoRI
PstI
DraIII
XmnI
AhaII
XhoI
HindIII
EcoRI
PstI
EagI
amp
lacZ
4000
pUR2440
1000
4807 bps
3000
2000
pINU1
lacZ
XhoI
EcoRI
EcoRI
SacI
EagI
PstI
NcoI
DraIII

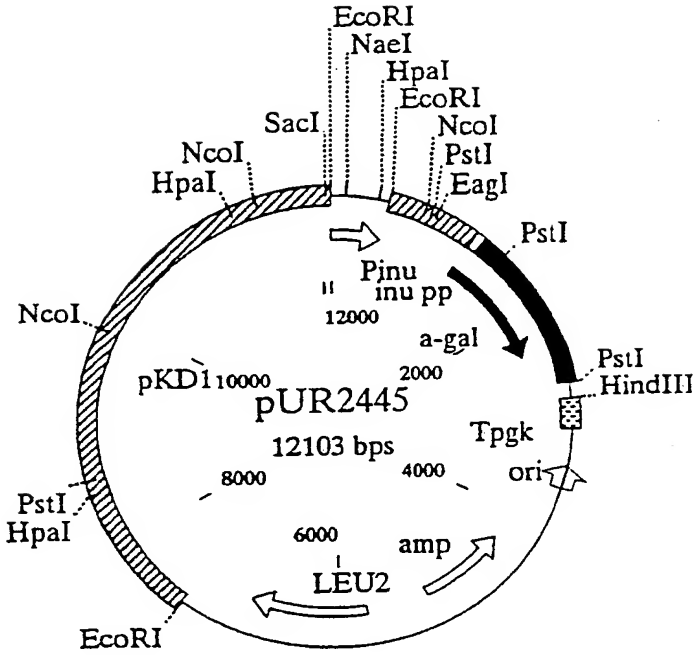
Figure 16.

1	GAATTCAGAG	CACAAGAATG	TGGAATCCCC	AGCTCTAGCA	AATGTCAGGT
51	GGAGACGGAC	ACACTTGTTA	CCATTCGTTA	CCAATTGTTA	ACACTACTTT
101	GCCTTCAACC	GTTCTTTCCC	CTGTGGGAGT	GCCTAGGCTT	ACCTTGGGAT
151	TGCCTAGACG	GAGACGCCTG	CGAGCGATTC	CTGAGTTATC	CCGTTCCCGT
201	GGCTACCCCA	AGTTTAATTC	TAATTCCAAT	TCCAATTCCA	ATTCTGATTC
251	CAATTCTGAT	TCCAGCTTGC	AACTTGCAAC	TTGCAACTTG	CAGTGGTAAC
301	CCCATCTGTG	CGCGCTCGTC	TGTGCATTTT	TCCCTTTTTT	TTCCGGCAGC
351	GCCGCCGGCC	CTTGTGTGCA	TAATTTAGCG	TTTTTGTGTT	TGTATTTTTT
401	TTTTTTTTTT	TTTGTGTTTG	CTTCTCTTGT	CGTTTTTGTG	TCTGTTTACC
451	TTTTCGGAAT	TGAGGCGTTT	CTTTTTGGCG	AATTC	

Figure 17: Sequencing primer INUT

5' ACG CAC TAA TTG ACT CAA CC 3'

Figure 18.


EcoRI
NaeI
HpaI
EcoRI
NcoI
PstI
EagI
SacI
NcoI
HpaI
PstI
Pinu
inu pp
12000
a-gal
NcoI
pKD1
10000
2000
PstI
HindIII
pUR2445
12103 bps
Tpgk
8000
4000
ori
PstI
HpaI
6000
amp
LEU2
EcoRI

A. CLASSIFICATION OF SUBJECT MATTER
IPC 5 C12N15/81 C12N15/55 C12N15/56 C12N15/11 C12N15/62
C12N1/19 //(C12N1/19,C12R1:645)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,91 00920 (UNILEVER PLC & UNILEVER NV) 24 January 1991 see page 6, line 47 - page 7, line 36	1,5,7
Y	see example 14 see claims 1-4,8,16-18,23-25 ---	6-8
Y	CURRENT GENETICS vol. 21, no. 4-5 , April 1992 , BERLIN, D pages 365 - 370 BERGKAMP, R. ET AL. 'Multiple-copy integration of the alpha-galactosidase gene from Cyamopsis tetragonoloba into the ribosomal DNA of Kluyveromyces lactis' cited in the application see the whole document --- -/--	6-8

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 March 1994	28. 04. 94

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Andres, S

Form PCT/ISA/210 (second sheet) (July 1992)

International Application No
PCT/EP 93/03547

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indi[redacted]where appropriate, of the relevant passages[redacted]	Relevant to claim No.
A	APPLIED AND ENVIRONMENTAL MICROBIOLOGY vol. 56, no. 11 , November 1990 , AMERICAN SOCIETY FOR MICROBIOLOGY pages 3337 - 3345 ROUWENHORST, R. ET AL. 'Structure and properties o fthe extracellular inulinase of Kluyveromyces marxianus CBS 6556' cited in the application ---	
P,X	APPLIED MICROBIOLOGY AND BIOTECHNOLOGY 40 (2-3). 1993. 309-317 Bergkamp R J M et al 'Expression of an alpha-galactosidase gene under control of the homologous inulinase promoter in Kluyveromyces' see the whole document -----	1-4,6-8, 10

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 93/03547

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9100920	24-01-91	CA-A-	2063592	08-01-91
		EP-A-	0481008	22-04-92
		JP-T-	5501949	15-04-93